DOCUMENT RESUME

ED 180 312 HE 012 037

AUTHOR Ericksen, Stanford C., Ed.: Cook, John A., Ed.
TITLE Support for Teaching at Major Universities.
INSTITUTION Committee on Institutional Cooperation.
PUB DATE 79
NOTE 114p.
AVAILABLE FROM Center for Research on Learning and Teaching, 109 E. Madison, Ann Arbor, MI 48104

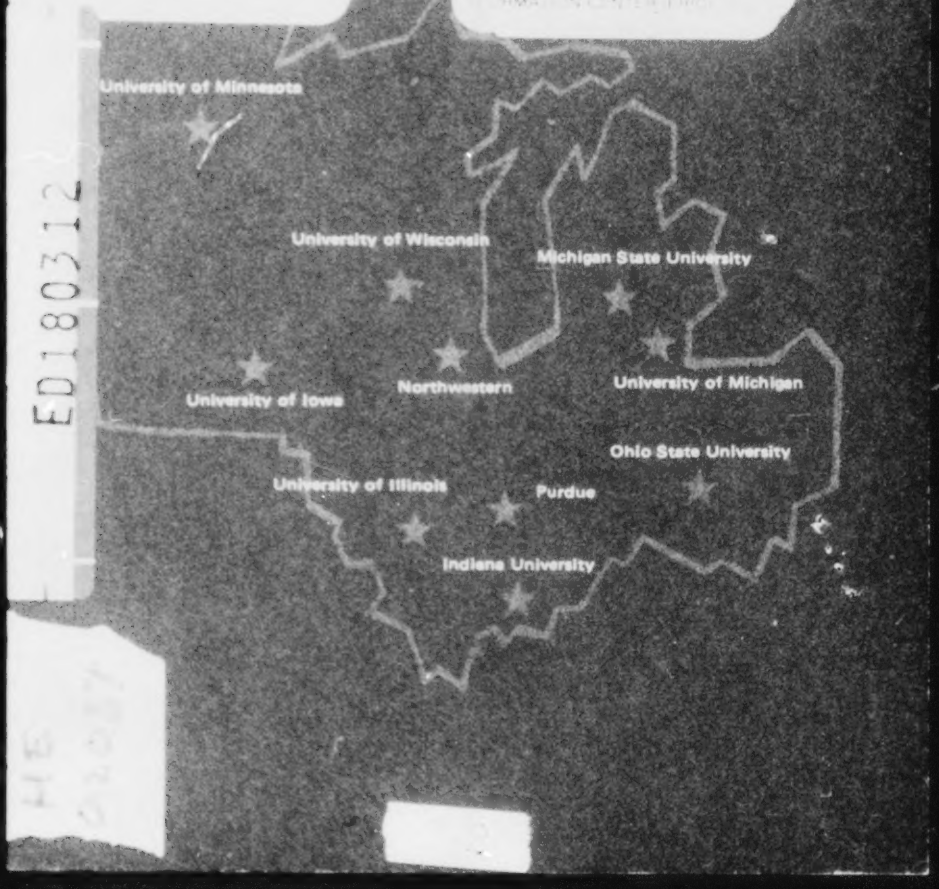
EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *College Cooperation: *College Faculty: Consultation Programs: *Educational Assessment: Educational Finance: Educational Technology: *Faculty Development: Financial Support: Grading: Higher Education: Information Dissemination: Institutional Evaluation: *Instructional Improvement: Interinstitutional Cooperation: Program Development: *Resource Allocations: Student Evaluation of Teacher Performance: Student Role: Student Testing: Teacher Seminars: Teacher Workshops: Universities
IDENTIFIERS *Teaching (Occupation)

ABSTRACT

Major themes that dominate institutional support for teaching are addressed by representatives of large research-oriented institutions of higher education that form the "Panel on Research and Development of Instructional Resources." The Panel is composed of the Big Ten universities and the University of Chicago. Contents are as follows: "The Panel and Its Purpose," by Stanford C. Ericksen, University of Michigan: "Faculty Development," by B. Claude Mathis, Northwestern University: "Providing Information," by Stanford C. Ericksen: "Special Funds for the Improvement of Instruction," by Robert H. Davis, Michigan State University: "Instructional Technology," by Gene K. Faris, Indiana University: "Seminars, Workshops, and Consultation Services for Faculty," by Russell W. Burris, University of Minnesota: "Testing and Grading," by Lowell F. Schoer, University of Iowa: "Student Evaluations of Instruction," by Warren P. Seibert, Purdue University: "Program Review and Instructional Development," by Robert G. Arns and William Poland, Ohio State University: and "The Systematic Involvement of Students in Program Evaluation," by Charles J. McIntyre, University of Illinois. The University of Wisconsin was unable to contribute a chapter. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Support for Teaching at Major Universities



ED180312

The Committee on Institutional Cooperation (CIC) was established in 1959 by the Big Ten Universities and the University of Chicago to facilitate pooling of resources and talents on matters of mutual concern in higher education.

Support for Teaching at Major Universities

Edited by

Stanford C. Ericksen

with

John A. Cook

Fall 1979

Written and distributed
by The CIC Panel on Research
and Development of Instructional
Resources

Published by
The Center for Research on Learning and Teaching
The University of Michigan
Ann Arbor, Michigan 48109

CONTENTS

CHAPTER ONE:	The Panel and Its Purpose Stanford C. Ericksen University of Michigan	5
CHAPTER TWO:	Faculty Development B. Claude Mathis Northwestern University	9
CHAPTER THREE:	Providing Information Stanford C. Ericksen University of Michigan	20
CHAPTER FOUR:	Special Funds for the Improvement of Instruction Robert H. Davis Michigan State University	32
CHAPTER FIVE:	Instructional Technology Gene K. Faris Indiana University	42
CHAPTER SIX:	Seminars, Workshops, and Consultation Services for Faculty Russell W. Burris University of Minnesota	51
CHAPTER SEVEN:	Testing and Grading Lowell F. Schoer University of Iowa	59
CHAPTER EIGHT:	Student Evaluations of Instruction Warren F. Seibert Purdue University	68
CHAPTER NINE:	Program Review and Instructional Development Robert G. Arns & William Poland Ohio State University	89
CHAPTER TEN:	The Systematic Involvement of Students in Program Evaluation Charles J. McIntyre University of Illinois	104

CHAPTER ONE

THE PANEL AND ITS PURPOSE

Stanford C. Ericksen
The University of Michigan

The Big Ten universities are usually considered "large research-oriented institutions," but they also are deeply involved with teaching. This report is an analysis of how good universities sustain and strengthen teaching.

In the spring of 1964, the directors of three facilities for research on learning and teaching (Charles McIntyre--University of Illinois, Russell Burris--University of Minnesota, and myself--from the University of Michigan) met informally to discuss the possibility of setting up an information exchange about teaching practices in the Big Ten universities. This led to the "Panel on Research and Development of Instructional Resources" within the structure of the Committee on Institutional Cooperation (CIC)--a long-standing organization of the Big Ten universities and the University of Chicago.

Each university named a representative to the Panel, which has met once or twice each year for the past 15 years. As individuals we engage in a constant flow of telephone calls and correspondence; we have come to know each other well. I serve as chairman of the Panel and have developed a genuine attachment to my counterpart "specialists." Each of us concentrates on matters related to teaching, and we have confidence in our abilities to identify significant factors pertaining to institutional support for teaching. The Panel is in close touch with the realities of undergraduate and professional-school instruction.

In addition to talking with one another, we publish an annual compendium of successful instructional projects from the CIC institutions--Development and Experiment in College Teaching. Each school provides up to four examples of interesting projects for which there is some evaluative information. One-page abstracts are

indexed under two major categories: "Subject-Matter Area" or "Institutional Support." Each report also includes a cumulative index. Most of the discipline areas are represented in these compendia, and the reported projects sample the more significant changes in instructional arrangements characteristic of college-level instruction. (For information about obtaining the reports, write to: The Center for Research on Learning and Teaching, 109 E. Madison, Ann Arbor, MI 48109.)

Modifying our roles as reporters, we on the Panel decided to take a self-critical look at some of the more pervasive themes that dominate institutional support for teaching. We are writing for an audience of university administrators, department chairpersons, educational research and development specialists, and university professors who are particularly interested in conceptual issues surrounding their more specific activities as classroom teachers. Each author has had ten or more years experience in educational research, administration, and teaching. Each is writing on a subject with which he has had considerable involvement in the setting of a large university. Nine of the Big Ten institutions are represented in this book. (Due to other commitments, the panel member from the University of Wisconsin was unable to contribute a chapter.)

A little should be said about the selection of topics. Because none of us is especially knowledgeable about the counseling functions of the college teacher, for example, this topic was left out. Similarly, we omit chapters on classroom dynamics and remedial instruction. These omissions represent topics that may be important to a particular teacher in a particular classroom; in planning this book, however, the group decided that greater attention should be given to the more general factors relating to institutional support for teaching.

In Chapter Two, "Faculty Development," Claude Mathis (Northwestern University) makes it clear that university professors do not stand still; they go to conventions, carry books and journals home to read, pursue scholarly goals, and change their ways of teaching in the classroom. They are not interested in "tinkering with teaching" and mistrust the label "faculty development." Mathis describes the organizational arrangements by which the Big Ten schools maintain continuing support for teachers.

As author of Chapter Three, "Providing Information," I take my cue from the attachment professors feel to the library. They treasure having access to information they can use for their professional development as scholars and teachers. Unfortunately, the dissemination mechanisms for teaching lag far behind the well developed networks for reporting research. This chapter summarizes various means for giving faculty ready access to information about teaching.

In Chapter Four, "Special Funds for the Improvement of Instruction," Robert Davis (Michigan State University) gives a perspective on how to allocate money to improve teaching. Specifically, he asks how special funds can best be used to improve

instruction. The answers to this question gain in significance as funds get tighter and the complexities of college teaching increase. Davis points out that money alone is not an adequate solution.

Stepping up the use of media is often a convenient and visible means of showing that one is an "innovative" teacher. In Chapter Five, "Instructional Technology," Gene Faris (Indiana University) explains how the technological resources at an institution can best be used to serve the instructional purposes of the faculty.

When a college professor takes the time to attend a workshop or seminar, will she leave with the feeling that her time was well spent? In Chapter Six, Russell Burris (University of Minnesota) makes it clear that in a workshop, the how-to-do-it should not be separated from the why-to-do-it.

In Chapter Seven, "Grading and Testing," Lowell Schoer (University of Iowa) discusses the varieties of grading systems and testing used in the Big Ten schools. Evaluation is a serious responsibility for teachers and frequently calls for institutional support.

The evaluation of teachers by students also concerns the Big Ten schools. Teachers have always been sensitive to signals from their students; now the use of student ratings is being formalized within institutions. In Chapter Eight, Warren Seibert (Purdue University) raises important questions about "Student Evaluations of Instruction."

Our book closes with two chapters on a persistent problem: evaluating the quality of educational programs. In Chapter Nine, "Program Review and Instructional Development," Robert Arns* and William Poland (Ohio State University) present the scope and some of the complexities of carrying out program review. They give special emphasis to the impact of instruction. In Chapter Ten, Charles McIntyre (University of Illinois) describes a specific procedure developed at his university to compile the evaluative judgments of students. Both these chapters reflect the kind of professional competence and educational leadership required to support instruction. It is one thing for a department chairperson, a dean, a vice-president, or a president to give a subjective evaluation of a given instructional/educational program; it is quite another matter to gather evaluative information as part of the collegial governing process. Accountability is not a popular concept--especially not as it moves from the student to the teacher, the department, the college, the university, and in some instances to the state-wide system of higher education. Program evaluation involves everyone from top to bottom, and institutions must be sophisticated about the technical procedures most likely to produce relevant information to support policy decisions and constructive action. Program evaluation is a good example of a "leading edge" development.

---

*Dr. Arns is now Vice President for Academic Affairs at the University of Vermont.

The substance of these chapters is not limited to large universities, although we recognize that qualitative modifications will be necessary when applying these analyses in smaller liberal arts or community colleges. We are confident, nonetheless, that the findings and the generalizations developed in the following pages will be useful wherever institutions are making a special effort to improve learning and teaching.

CHAPTER TWO

FACULTY DEVELOPMENT

B. Claude Mathis
Northwestern University

Faculty development appeared on the academic scene during the 1970's with a suddenness which suggested that ultimate and satisfying solutions for the ills of higher education were finally at hand. Higher education's preoccupation with the effective use of resources has stimulated new approaches to solving old problems. This "new wine" has, at least for the moment, become faculty development, a label for the recognition that faculty members themselves are the answers for the major problems of the post-secondary institutions they serve. Like so many movements in education, faculty development has emerged as an amorphous force for change without the benefit of concise definitions and meanings beyond the structure of the efforts which carry the label.

The Matter of Definition

Faculty development presumes that the quality of educational programs is directly related to the behavior of faculty, especially their effectiveness as teachers. Much of the literature about faculty development implies that a new approach has been discovered to bring faculty back to teaching, in spirit if not in person. Some see it as a way of packaging more traditional methods for improving education so that a broader range of disciplines may be caught in the net of educational change.

The renewal of interest in faculty as teachers and educators, as contrasted to faculty as scholars and researchers, was initiated nationally by the Project to Improve College Teaching, jointly sponsored by the American Association of University Professors (AAUP) and the Association of American Colleges (AAC). Prior to the AAUP-AAC efforts, the federal government had established, through Title VI, Part B of the Higher Education Act of 1965, a granting mechanism for improving instruction in colleges and universities. The AAUP-AAC project was directed by Kenneth Eble, and it

produced two publications, one about career development and college teaching (Eble, 1971b), and the other about the evaluation of teaching (1971a). Both of these monographs received national attention and were followed by a book from Eble, Professors as Teachers (1972), which helped to call attention to the fact that the traditional role of a faculty member has always been that of the teacher-scholar committed to excellence in teaching as well as in research and scholarship. He argues persuasively for the rediscovery of this tradition.

The AAUP-AAC activities and reports were but a prelude to numerous publications which attempted to provide a rationale and a plan for change and renewal in higher education through the use of faculty development strategies and programs. However, as with much of what we commit ourselves to in our educational institutions, activities were begun far ahead of any demonstrable knowledge that faculty development strategies have impacts which justify their existence. Only the future can answer the question of whether faculty development is an educational fad whose interest has peaked or whether it is a strategy for change that will survive the excitement that created it.

How do those who contribute to the burgeoning literature on faculty development define the concept? Gaff, in his book Toward Faculty Renewal (1975), has this to say:

> Faculty development, although not a new term, assumed a new sense of importance during the early 1970's . . . It may be defined as enhancing the talents, expanding the interests, improving the competence, and otherwise facilitating the professional and personal growth of faculty members, particularly in their roles as instructors. (p. 14)

In its contemporary usage, faculty development first attracted national attention in the title of a monograph from Change Magazine, " Faculty development in a time of retrenchment," authored by the Group for Human Development in Higher Education (1974). While not defining the term as explicitly as Gaff, the thrust of the publication linked faculty development with a retrenchment mode in higher education and called for a number of reforms which would elevate the importance of teaching to a position of parity with research and scholarship.

One of the most frequently cited articles which attempts to define faculty development is by Bergquist and Phillips (1975). The abstract at the beginning of their statement is a summary of their point of view:

> Since piecemeal efforts to improve college and university teaching have generally proven ineffective, we must turn to a comprehensive approach to faculty development, through which we can develop new methods of evaluation and diagnosis, find viable ways of introducing new technology and curricula, and explore new approaches to instructional improvement. Faculty development must give serious attention to the impact of change

> on the faculty member himself and on his institution. Organizational and personal development thus become essential to faculty development. It is only through such a comprehensive approach that efforts toward improvement can have lasting impact. (p. 177)

Bergquist and Phillips believe that faculty development efforts must have an impact on change in the attitudes of faculty, on change in the process of teaching, and on change in the structure of the institution. Efforts which attempt to change only one of these variables without answering the implications for the other two do not qualify as faculty development activities. Organizational, instructional, and personal contingencies determine the comprehensive nature of faculty development programs. Without this comprehensive commitment, a faculty development program would be misnamed.

Other writers have contributed to a general meaning for faculty development (Francis, 1975; Mathis, 1974; Simerly, 1977; Toombs, 1975; Wergin, Mason, & Munson, 1976). These sources emphasize activities, incentives, support structures, and policies which help faculty members to examine and improve the educational process. The comprehensive nature of faculty development includes all the roles and responsibilities of faculty, not just teaching, although the improvement of the teaching/learning alliance is the major goal of faculty development programs.

The ideology of the faculty development movement has a number of identifiable relationships to a prior scholarship in higher education. The human relations movement, represented most prominently by the National Training Laboratories, has been a major influence on the commitment of faculty development programs to the personal development of faculty, especially as personal factors relate to professional roles. A maturity in interpersonal relationships should transfer to specific professional relationships. Thus, to a large extent faculty development represents a commitment to improving the educational process by concentrating on an understanding of self as the point of entry to the examination of teaching and learning. The faculty member's own needs, values, and attitudes become the central concern--rather than specific methods of teaching to which faculty should conform.

Faculty development is interlaced with the study of strategies for change in higher education, which is itself part of the human relations movement in education. Faculty become the agents of change who have the power which, if successfully applied, should result in more equitable incentive systems, greater value being given to effective teaching, and the utilization of faculty resources for improvement of the quality of education rather than for professional self-aggrandizement.

Faculty development is a manifestation in higher education of the recent interest in adult human development which has emerged during the past decade. While the interest itself is contemporary, the literature on adult development has a respectable history. The

work of Sanford at the Wright Institute (1971) and statements by Hodgkinson (1974) and, in a more popular vein, by Sheehy (1976) have drawn attention to such concerns as professional socialization, mid-career crises, and critical transition points in academic careers.

Faculty development has also grown out of a crisis mentality in higher education dominated by such words as reexamination, renewal, retrenchment, and revitalization, among others. The use of such language suggests that faculty development is an attempt to minimize stress by improving the communication networks among faculty. Effective teaching represents one common denominator across disciplines.

The field of scholarship known as higher education as well as centers for the study of higher education have been a central influence in shaping the literature about, and in designing programs for, faculty development. This is in contrast to other approaches to the improvement of teaching, especially instructional development, which have had their genesis in the behavioral sciences, especially psychology. Faculty development is not a movement dedicated so much to experimentation as it is to planned change. It avoids the dominance of an empirical or experimental methodology, and instead relies on a rationalism which is both analytical and intuitive.

Gaff (1975) adds to our understanding of faculty development in the following statement comparing faculty and instructional development:

> Faculty-development programs . . . focus on faculty members rather than on the courses they teach or their organizational environments; they stress both professional and personal development of professors over the life system, stemming from their intellectual roots in developmental, clinical, and social psychology and psychiatry. In contrast, instructional-development programs focus on the design of effective conditions of learning, particularly as provided by courses, emanating from their roots in the fields of curriculum and instruction, learning theory, educational media and technology, and systems theory (p. 174).

Although the intellectual roots may be different for these two approaches to improving teaching, faculty development and instructional development differ only by a matter of degree. If anything, instructional development is one facet of the comprehensive nature of faculty development, according to Bergquist and Phillips (1975).

Steps toward Implementation

Faculty development programs have appeared nationally more frequently in those colleges whose major educational programs are for undergraduates and whose ethos values teaching as a primary commitment of faculty. One of the outstanding successes has been in

the institutions which belong to the Council for the Advancement of Small Colleges (CASC), a consortium of small liberal arts colleges, many of which are church related. CASC has been able to attract considerable support from foundations and other sources to use in helping member institutions plan and implement faculty development programs designed to address the special needs of each college. CASC staff are able to provide resources which may not be available on the home campus and bring colleges with like problems together to seek cooperative solutions. A handbook which has received wide use outside of CASC membership is available to help colleges plan faculty development programs (Berguist, Phillips, & Quehl, 1975, 1977).

Faculty development initiatives, at least those labeled as such, have not appeared with any frequency in large research-oriented universities. In those instances where major universities have used the term "faculty development" to label improvement programs, the designation has been primarily a response to the availability of foundation funds for faculty development. This is not to say that universities are resistant to changes which might enhance the quality of educational programs. Most universities are sensitive to the need for effective teaching by faculty, but they approach the problem through strategies which are better classified as instructional development. A recent report on efforts to improve teaching through campus centers at nine major universities in the United States identifies four primary themes which are a better fit for instructional development strategies (Mathis, 1978). These objectives are:

(1) Activities are primarily designed to serve the educational interests of undergraduates, and undergraduate courses are given primacy.

(2) Instructional innovation through the use of educational technology, media, and specific teaching methods is a dominant program feature.

(3) The improvement of undergraduate instruction has much time devoted to increasing faculty awareness about teaching.

(4) A high priority is given to the evalution of instruction, especially through student ratings.

Nowhere is the status of faculty development more equivocal than in the institutions which make up the membership of the Committee on Institutional Cooperation, those midwestern universities known as the Big Ten. None of these institutions has a comprehensive program for the improvement of education which is called faculty development, yet they all engage in some faculty development efforts under other labels. Most of these institutions have university-wide programs to improve teaching, and most have offered for over a decade many of the activities which are being viewed as unique to the faculty development movement. One university, Northwestern, has a center which is identified with

faculty development activities nationally to a greater degree than other Big Ten programs, yet the impact on its own campus has been less than its impact nationally.

Many reasons can be advanced for the ambivalence which Big Ten institutions feel about faculty development. The major source of this ambivalence can be found in the nature of the university itself. Teaching in a university responds to forces unique to universities:

(1) Rewards within the university: Although changes have taken place during the past decade, the primary research system emphasizes competence in scholarship and research first, and effectiveness in teaching second. Advancement in Big Ten institutions is easier if one is an effective researcher rather than an effective teacher. Effectiveness in both research and teaching is the goal, but research still has primacy.

(2) Graduate students and their involvement in teaching: Teaching assistants provide a significant portion of undergraduate instruction, especially at the freshman and sophomore level in most Big Ten universities. Teaching assistants, for the most part, are involved in the pursuit of a Ph.D. Their commitment to effective teaching is often secondary to the successful completion of their degrees. As transient teachers, their individual effects on the quality of teaching are short-term impacts. The liberal use of TAs gives faculty the time they need to do their research, and the research often provides the money necessary to "buy one's way" out of the classroom by supporting a graduate student who becomes a TA. This system does not include the incentives which would help to make the TA a dedicated teacher.

(3) The existence of many faculties rather than one faculty: Big Ten universities all have a primary commitment to professional training. Many of their professional schools are among the best in the nation; however, this excellence leads to a diversity of values and attitudes which result in faculty roles being defined more by identification with one's discipline than one's institution. This wide range of disciplines and professions does not lead to a simple consensus about the value of teaching or the best way to teach.

These three characteristics of universities present problems which are amenable to faculty development solutions, but in guises which omit some of the characteristics of faculty development discussed at the beginning of this paper. For one thing, research-oriented faculty members such as those who populate Big Ten institutions are skeptical of programs to improve teaching which get too "personal." The personal dimension of faculty development and its dependence on human relations training, encounter techniques, and group strategies for increasing awareness is not the kind of empiricism most Big Ten faculty members trust.

The wide diversity of values and attitudes which faculty have about teaching results from the influence of their disciplines and their identification with various professions. This diversity makes it difficult to implement a general faculty development program throughout the total university and expect representative participation from all faculty groups. The professor of chemistry, engineering, or dentistry approaches teaching differently than the professor of English, social work, or Slavic languages. These differences suggest that faculty development in Big Ten universities could best emerge from within their departments and professional schools rather than from centers and programs which serve the total community; although as emphasized in the chapters by Faris and Davis, providing information and special funds can best be implemented through a central unit.

The reward structure in Big Ten institutions is only beginning to respond to effective teaching as a behavior to be rewarded. Most faculty still view research and scholarship as the primary determinants of promotions, tenure, and increases in salary. Effective teaching does not have parity with research and scholarship in influencing the distribution of rewards in Big Ten schools.

Big Ten universities have not been victims of the retrenchment ideology which has severely threatened other institutions. If the faculty development movement is a response to crisis, then Big Ten institutions have not as yet had a severe enough crisis to warrant comprehensive faculty development as a solution. Salaries are respectable within the CIC schools; tenure has not eroded; student enrollments are not declining precipitously; and endowments and state support are sufficient to allow an orderly retreat if things get worse.

Excellent examples of support for faculty development and faculty involvement in the examination and improvement of teaching and learning are found in CIC institutions. The Center for Research on Learning and Teaching at the University of Michigan, the Instructional Development and Telecommunication Services at Michigan State University, the Office of Instructional Resources at the University of Illinois, the Consulting Group on Instructional Design at the University of Minnesota, the Learning Resources Programs within the Indiana University system, and the Center for Teaching Professions at Northwestern University all are university-wide centers or programs offering a comprehensive range of activites for improving teaching and learning. Their clients are faculty members and teaching assistants. While their basic approaches best fit the values of each institution, generally these approaches view teaching as a behavior which each faculty member must manage for himself or herself, independent of a prescription or formula. All of these centers and programs recognize that disciplines and professions have differing perceptions of education, and this recognition is reflected in varying options which are available for working with faculty. Each program emphasizes individual consultation with faculty rather than group-based interventions.

These six programs, as well as others in CIC universities, have been involved in shaping institutional policies about the importance of rewarding effective teaching, and providing support to sustain it. The primary mission of these units, however, is to improve the process of education by working with faculty who teach. The personal dimension of faculty development is not a major concern, nor is the organizational dimension.

The Center for Research on Learning and Teaching (CRLT) at the University of Michigan was established in 1962 on the recommendation of a faculty committee. As an agency of the faculty with a broad and flexible range of research and development rsponsibilities, CRLT has made a commitment to research, development, evaluation, and dissemination as major components of its program. These four objectives represent positive values within the University of Michigan and have given CRLT credibility as something other than a service center populated by technicians. The principle dissemination device is its Memo to the Faculty, which has become a classic example of how to communicate creatively about teaching to scholar-teachers who want facts and evidence rather than patronizing or platitudinous advice. The Memo has been used as a basic source of information in faculty development programs nationally.

The first director of CRLT, Dr. Stanford Ericksen, has chaired the CIC Panel on Research and Development of Instructional Resources since it started in 1964 and has been the major force behind the publication of Development and Experiment in College Teaching, a yearly report on instructional and educational projects and programs at Big Ten universities and the University of Chicago. It is interesting to note that faculty development first appeared as a general index topic in that publication in the issue for 1973. Many items in previous issues, however, were reclassified as faculty development in the section listing university-wide activites (CIC, 1966-1979).

The Instructional Development and Telecommunication Services Center at Michigan State University represents a number of programs as part of a comprehensive attempt to improve the quality of teaching and learning at MSU. These efforts have been among the most successful nationally. The large investment which Michigan State University has made in providing a technological base for instruction gives the appearance of an almost total concentration on instructional development activities. Such is not the case, however. Educational programs at MSU are based on recognition of the importance of instructional, personal, and organizational components in generating effective-teaching. A recent report on these programs defines faculty development as "activities performed to promote individual growth and development of teaching personnel . . . most faculty development activities concentrate on an exploration of attitudes about teaching and learning and acquisition of skills and knowledge related to teaching" (Davis, Abedore, & Witt, 1976, p. viii). Michigan State University, especially through the activities of the Educational Development Program, which provides teaching improvement grants to faculty, and the Learning and

Evaluation Service, which offers individual consultation for faculty, has a wide range of faculty development activities for the improvement of teaching.

The Consulting Group on Instructional Design at the University of Minnesota is an all-university unit which links research about human learning with implementations in the educational programs of the university. This "linkage" function is especially appropriate for a faculty which values a research base for educational efforts. Successful projects in art history and legal education have demonstrated the sensitivities of those in the Consulting Group to the potency of different points of view on teaching and learning, and to the impact varying values can have on instructional design.

Indiana University's faculty development efforts include activities within the eight-campus system and were initiated in the Audio Visual Center on the Bloomington campus in 1967-68 with the assistance of a grant under Title VI, Part B of the Higher Educational Act of 1965. The four-year funding cycle under this act enabled the University to create a faculty development program in the Office of the University Academic Vice President (presently in the President's Office) that has expanded its activities each year. In addition to this centralized effort, various kinds of faculty development activities have been undertaken on each campus. Examples of these activities are: a four-year program supported by the Lilly Endowment and conducted jointly by the Office for Learning Resources and academic departments on the Bloomington campus, designed to establish a dialogue between faculty members on the Bloomington campus and faculty members in four-year institutions concerning the content of their respective disciplines; a teaching consultation process which has been employed on an intensive basis at the Indiana University Northwest campus and less intensively at other campuses, to allow the faculty member a comprehensive and in-depth look at classroom instruction; and University-wide activities which include the individual analysis of teaching performance, workshops on the preparation of dossiers for promotion and tenure, the funding of summer faculty fellowships for the improvement of teaching, and the coordination of distinguished teaching awards and the appointment of distinguished professors.

The Center for the Teaching Professions at Northwestern University has evolved its efforts to fit the demands of a private university which has neither a large investment in educational technology nor a faculty willing to stray too far from traditional methods in their teaching. The Center has been involved in a program of faculty development for the Associated Colleges of the Midwest, a group of thirteen private liberal arts colleges; Center staff have concentrated their efforts on solutions which seem more appropriate for faculty in the private sector of higher education. The criteria most frequently used by Center staff to assess effectiveness are those which indicate movement in faculty from dependence to independence (or self-direction) in the analysis, understanding, and control of factors which influence a positive and satisfying feeling about what one does, and is able to do, as a teacher-scholar in the context of a specific institutional culture.

Some prominent features of the programs of the Center for the Teaching Professions are: a "Seminar on College Teaching" for graduate students whose career goals include teaching--those taking the seminar obtain credit toward their degrees; a Visiting Scholars program which brings faculty from other institutions to the Center for varying periods of time; and a "faculty-teaching-faculty" program which has resulted in several books and monographs about teaching.

Many more examples of specific faculty development activities are available from other CIC institutions and are described elsewhere in this report. Whether or not faculty development will emerge in CIC universities as an identifiable strategy with a scholarship to support it will depend on the needs of faculty at those institutions for whatever faculty development has to offer over and above the programs now available. Much also will depend on whether or not faculty development emerges as a concept which can be defended as something different and/or new in higher education. Perhaps faculty development is an attitude or a "belief-value matrix" which represents only old wine in a new bottle. Certainly the concept is in need of an empirical scholarship all its own to defend its supposed uniqueness. Until that is established, Big Ten faculty will continue to be ambivalent about the term, and their universities will continue to offer faculty development opportunities under other labels.

REFERENCES

Berquist, W. H., & Phillips, S. R. Components of an effective faculty development program. Journal of Higher Education, 1975a, 46(2), 177-211.

Berquist, W. H., Phillips, S. R., & Quehl, G. H. (Eds.). A handbook for faculty development (Vol. 1). Washington, D. C.: Council for the Advancement of Small Colleges, 1975.

Berquist, W. H., Phillips, S. R., & Quehl, G. H. (Eds.). A handbook for faculty development (Vol. 2). Washington, D. C.: Council for the Advancement of Small Colleges, 1977.

Committee on Institutional Cooperation--Panel on Research and Development of Instructional Resources. Development and experiment in college teaching (Reports #1-14). Ann Arbor, MI: Center for Research on Learning and Teaching, University of Michigan, 1966-1979.

Davis, R. H., Abedore, A. J., & Witt, P. W. F. Commitment to excellence: A case study of educational innovation. East Lansing, MI: Michigan State University, 1976.

Eble, K. E. The recognition and evaluation of teaching. Washington, D. C.: American Association of University Professors, 1971a.

Eble, K. E. Career development of the effective college teacher. Washington, D. C.: American Association of University Professors, 1971b.

Eble, K. E. Professors as teachers. San Francisco: Jossey-Bass, 1972.

Francis, J. B. How do we get there from here? Program design for faculty development. Journal of Higher Education, 1975, 44(6), 719-732.

Gaff, J. G. Toward faculty renewal. San Francisco: Jossey-Bass, 1975.

The Group for Human Development in Higher Education. Faculty development in a time of retrenchment. Change Magazine, 1974.

Hodgkinson, H. Adult development: Implications for faculty and administrators. Educational Record, 1974, 55, 263-274.

Mathis, B. C. Persuading the institution to experiment: Strategies for induction (Occasional paper #9). Evanston, IL: Center for the Teaching Professions, Northwestern University, 1974.

Mathis, B. C. The university center. In J. Lindquist (Ed.), Designing teaching improvement programs. Berkeley, CA: Pacific Soundings Press, 1978.

Sanford, N. Academic culture and the teacher's development. Soundings, Winter 1971, pp. 357-371.

Sheehy, G. Passages. New York: E. P. Dutton & Co., Inc., 1976.

Simerly, R. Ways to view faculty development. Educational Technology, 1977, 17(2), 47-49.

Toombs, W. A three-dimensional view of faculty development. Journal of Higher Education, 1975, 44(6), 701-718.

Wergin, J. R., Mason, E. J., & Munson, P. J. The practice of faculty development. Journal of Higher Education, 1976, 47(3), 289-307.

CHAPTER THREE

PROVIDING INFORMATION

Stanford C. Ericksen
University of Michigan

It may be in order to tell things to students, but don't "lecture" the faculty on how to teach. Information about teaching should be available, not imposed. Each teacher is a distinctive person who will inevitably interpret and adapt information in a way consistent with his or her unique style of teaching and in terms of the requirements of a given course and its students. With information about teaching, we face the educational version of the universal problem of how to apply general knowledge in a specific setting.

College professors prefer to think of themselves as their own best "doctors" in the diagnosis and treatment of their teaching problems. A distinctive feature of the college culture is that faculty members exercise their autonomy as individuals, making their own decisions about what and how to teach. This academic freedom is also buffered by the tradition that "the classroom is my castle" or by the belief that the "art" of teaching does not lend itself to objective analysis and review. In taking these positions, teachers are not necessarily trying to construct a criterion-free task for themselves or to avoid evaluative judgments. Rather, they want to preserve those distinctly personal styles of teaching which they believe have an influence on students. The professor who says, for example, "My job is to give students the opportunity to observe the scholarly mind at work," emphasizes the teacher's role as model.

Another professor may reflect a quite different concern for teaching when she asks, "How do I develop a computer simulation of a conceptual problem in plant genetics?" Instruction is an omnibus activity which may include setting an example for students, giving a fact-filled lecture, making effective use of technological aids, or pursuing the analysis of value judgments growing out of a given discipline area. Within this multi-dimensional complex, teachers are alert to information they might use to improve their competence.

A good school will support this interest by making readily available worthwhile information about how students learn and how teachers influence this process.

Who Tells What To Whom?

Given the highly personal quality of teaching, what additional information may be needed and from what sources? The answer depends. The teacher who has just participated in a workshop on instructional technology may benefit from access to a media specialist. Another teacher fearing, "I don't come across with my students," may turn for assistance to a respected departmental colleague. In any case, very few college teachers are smug about the quality of their work; they sense what changes might be made in how they manage the classroom hour and are receptive to ideas picked up in casual conversations with friends, colleagues, students, and peers in their subject area. They also read about teaching, but the impact of the printed word is limited. Most written reports deal with general cases and broad ranging ideas which, to be useful, must be transformed by the reader within his or her idiosyncratic style of teaching.

The Small Group Interchange and Personal Consultation

Teaching is a common denominator among faculty members, but the agenda for the usual departmental staff meeting does not focus on teaching as a specific process, on such classroom problems as how to respond to expressions of hostility from students, or how to cope with indifference, boredom, or anxiety. Staff members are not often encouraged to share their frustrations or feelings of inadequacy--even less so, their successes. As a consequence, teachers sometimes feel isolated and thus appreciate the opportunity to discuss, for example, what they do as teachers, their own perceptions of the values of students, the reward structure of the university, or the changing pressures from the off-campus community. While it is important for teachers to gain direct experience with new modes of instruction--as in workshops--it may be more valuable simply to talk things over with colleagues who share similar interests, values, and problems.

A conversation between a teacher and a credible consultant may be the most direct arrangement for providing information about teaching. In its final set of recommendations for elevating the importance of teaching, representatives for a group of 11 colleges and universities stated, "Access to individual consultation is the most effective means for adapting instructional arrangements to the particular characteristics of each teacher" (Ericksen, Criteria VIII, 1978). Consultation is important because the teacher controls what is talked about, namely, his or her own problems and concerns. Russell Burris gives explicit accounts of effective consultation and workshop arrangements in Chapter Six.

Curricular and/or Instructional Change

For the purpose of this analysis, I specify three major functions of the teacher: to plan the content and organization of a course of study, to manage the classroom hour, and to make fair and discriminating judgements about student performance. The first two components are generally given most weight in the evaluation of a teacher's competence. (In principle, the third is equally important, since it is a functional measure of what the instructor believes important that the students carry away.) Inasmuch as the subject matter may be taught better by one teacher than by another, it is appropriate to distinguish curricular information from how a course is taught.

When discipline specialists get together to talk about teaching --in a departmental staff meeting or at a convention--their discussions usually concern course content rather than more personal topics such as how a teacher motivates or gets along with students or influences their understanding of subject matter. In the 1950's and early 1960's the National Science Foundation funded a number of educational commissions in physics, chemistry, geology, biology, etc. The publications and conferences sponsored by these groups were generally dominated by issues relating to curricular substance. Instruction was, for the most part, treated in terms of the use of various technological aids such as computers, films, TV, audiocassettes, etc.

This emphasis on subject matter and instructional technology reflects the emphasis these are given in the literature. The personal, ego-involved aspects of teaching do not lend themselves to public reporting. An interesting exception is made when a faculty colleague is selected for an "outstanding teacher" award. The citation is usually phrased in general terms reflecting the presumed qualities of the master teacher. The value of these awards as an institutional means for supporting teaching has been questioned and cynics refer to them as popularity polls. Nevertheless, this form of recognition could be an important source of information about the classroom manner of successful teachers.

What we need are more specific data about the basis for selection and how these are weighted in the process of peer evaluation of teaching. Is the "stimulus value" of a teacher in the classroom the only basis, or is careful appraisal given to the contributions away from the podium, such as preparing a significantly different course syllabus, making special adaptations for particular kinds of students, developing new resources for instruction (see Chapter Five on the media), providing unusually effective procedures for the evaluation of student achievement, and the like. To maintain the merit system of recognition and reward in a university under current conditions of limitations and stress will require explicit attention to the full range of a teacher's responsibilities.

The Higher Education System

The health and well-being of higher education as a social institution is a vital matter, and one which requires considerable attention in the literature. The Chronicle of Higher Education and Change Magazine are familiar sources of information about the larger aspects of higher education. The entire academic community is, of course, concerned about how the educational system is maintained. Fiscal support is obviously a central consideration: Is the tax burden equitable, is the education dollar well spent, where can economies be made, and how can significant new programs be budgeted and older units strengthened? The politics of higher education are also under constant review by faculties and administrations, and the Bulletin of AAUP regularly reminds us about problems of academic governance. The sociology of education is equally important as communities and states wrestle with problems of affirmative action, adapt to a changing profile of college students, and modify the curriculum in line with new vocational and cultural opportunities and demands.

Against this backdrop of frustrations and threats, who can afford to worry about the classroom interchange, the problems experienced by teachers working with students in given courses of study? Fortunately, many teachers do; they continue to ask questions about how best to organize a course and instruct their students. There is always room for improvement. Teachers know this, and in their private manner seek information--if not direct assistance.

Research and Service for Teachers

Most university teachers are keenly aware that their research and scholarly productivity, that is, their competence as discipline specialists, is given more recognition and reward by the home institution and by their discipline peers than is their performance as classroom teachers. Research is more important than teaching in the career progress of most members of the faculty in the Big Ten schools. Other types of schools, especially the liberal arts and the community colleges, give more explicit attention to teaching, and perhaps also to service contributions. In line with this institutional priority, large universities provide resources to sustain and to strengthen the competence of their faculty as leaders in the scholarly and scientific domains. The libraries are supplied; laboratories and research facilities are made available, as are grants to support research and scholarly work.

In contrast, teaching has more or less been expected to take care of itself--when teachers are given adequate salaries, classrooms, clinics, instructional laboratories, comfortable teaching loads, and clerical assistance. These expenses, of course, take up the major share of an institutional budget, and it is difficult for a school to favorably consider additional dollar demands in support of teaching. Nevertheless, many institutions have made special arrangements for giving teachers direct access to technical and professional assistance and for providing special funds for instructional development. These

resources are usually made available through a specially designated service/research unit on campus (see Chapter Two for a description of such units in the Big Ten universities).

The staff members in these facilities are prepared to answer questions which are, might, or should be asked by teachers and administrators. Teachers cannot always pull themselves up by their own bootstraps, so they seek assistance in many areas--individualized instruction, computers, media, simulation and gaming, role playing, experiential education, student development, motivation for learning, tutorial instruction, group discussion techniques, learning theory, the design of instructional projects, and procedures for the evaluation of students, teachers, and programs. The common feature in ths entire array is, of course, the transfer of information from a specialist to a faculty member whose own specialty lies more with course content than with the development of new procedures for teaching. Instructional specialty units are expected to help the faculty keep abreast of new developments and to assist them in managing the vast array of new resources on the college campus today.

Books, Journals, and Newsletters

Most college teachers learned about teaching by watching it being done. Conversations with peers over a cup of coffee, at cocktail parties, staff meetings, or seminars also combine to strengthen one's personal conception of the model teacher. (The local departmental culture is the primary source of information about how to get ahead as a teacher.) Reflections about one's self as a teacher are usually a rehearsal of one's successful ventures; we give greater attention to the satisfactions intrinsic in doing a good job teaching. From time to time, however, we admit to ourselves certain weaknesses or recognize instructional habits that need changing. We may even be moved to read about teaching, although books usually tell the reader more than he or she wants to know. Compared to scholarly reports in one's area, most books on teaching are "hammock" reading, such as Pirsig's Zen and the Art of Motorcycle Maintenance. Kenneth Eble's The Craft of Teaching is a perceptive analysis by another professor of English with a message relevant to the work of most college teachers. These books present a challenge but do not overly threaten the reader's relaxed satisfaction with himself as a teacher. The references at the end of this chapter list other recent books about college teaching.

I will not attempt to inventory the discipline-based journals that address matters of teaching. For many years articles in these journals tended to be essays or show-and-tell treatments of local projects. Two current trends are apparent: (1) Formal publications sponsored by professional organizations are giving more space to matters of teaching, and (2) more of these articles reflect empirical findings and conceptual developments related to instruction. The Journal of Medical Education, for example, frequently contains articles by educational research and development

specialists co-authored with their M.D. colleagues. Similarly, the Journal of Personalized Instruction makes an effort to report studies by teachers from various disciplines and with particular emphasis on evaluative information.

The year 1977 was the 25th year of publication of one of the more respected journals reporting research, new developments, and analysis in the general area of the media--AV Communication and Review, sponsored by the Association for Education, Communication, and Technology. A number of other journals in this field are sponsored by media industries, and their readers are mainly other media specialists. Computing technology is just getting started as an instructional aid and no single journal dominates this field.

A gap remains, namely, how to inform the full range of teachers about significant new developments in college teaching. The Educational Research and Information Clearinghouse (ERIC) is a nationwide bibliographic service giving particular attention to items that are not published in standard journals. A recent publication by Sally S. Gaff, Conrad Veste, and Jerry G. Gaff, Professional Development--A Guide to Resources (1978) is an excellent compendium of resources relevant to the various functions of a college teacher. The Change Magazine series "Report on Teaching" was a large-scale dissemination effort, and selected items have been combined in book form, Guide to Effective Teaching (1978). A distinctive feature of this program was the cooperation by several discipline organizations in identifying significant projects. Most of the reports were written by professional writers and distributed widely throughout American higher education. "Report on Teaching" contains minimum theory, with enough detail to give the reader a sufficient understanding of what went on and, presumably, can serve as a basis for converting the reported procedure for use in one's own course.

Newsletters come a dollar a dozen; their popularity indicates their usefulness. A local newsletter has the advantage of reporting here-and-now activities of fellow teachers, confirming that faculty colleagues are concerned about teaching and are doing something about it. As a rule, a teacher is more receptive to an analysis of teaching if it is tied closely to his or her discipline and to the home institution. Local reports about specific accomplishments in teaching reduce departmental provincialism and the isolation of the individual classroom teacher.

My own persistent involvement in "dissemination" is the Memo to the Faculty series published at our Center for Research on Learning and Teaching at The University of Michigan. The original idea for this newsletter was suggested by the Faculty Advisory Committee in the first year of the Center's existence (1962). Since this is a large university, the Committee felt that a guide would be useful in establishing and maintaining the visibility of the new unit as an agency by and for the faculty. These reports work both sides of the street--in application, but also in the context of the research and theory that justify a particular procedure. University professors may want to know "how to do it," but they also appreciate knowing the conceptual support that justifies a possible modification in their

more familiar modes of teaching. The last four Memos:

#60 - The Lecture
#61 - The Dimensions of Merit
#62 - The Discussion Period in the College Classroom
#63 - The Administration of Good Teaching

Three of the Big Ten universities (Illinois, Purdue, and Michigan) were involved in an eleven-school project to "elevate the importance of teaching." Over a period of two and a half years, the participants met for ten two-day sessions to review local activities, to analyze reports of programs elsewhere, and finally, to draw up a set of recommendations. A major goal of the project was to disseminate its findiags. This was mainly accomplished by way of a specially prepared newsletter: Criteria, For the Evaluation, Support, and Recognition of College Teachers. The last issue, Criteria VIII , "Recommendations for Elevating the Importance of Teaching," proposed three basic questions:

1. How can colleges and universities support faculty efforts to sustain or improve their teaching?
2. What are the most effective ways to recognize and reward teaching?
3. What are the key principles and procedures for evaluating teaching?

These questions are much easier to answer within the context of a single institution, but taking the local path will often bypass critical general issues that should be taken into account for the long haul. Most colleges and universities treasure their autonomy, but if due regard is not given to lessons learned at other schools about how to enhance teaching, the self-defining institution runs the risk of drifting into provincial ineffectiveness. The recommendations in the final issue of the Criteria reports set forth the general conditions, but these must be applied in different measure by each institution.

A somewhat different device for providing information is the annual compendium by the Panel of Research and Development of Instructional Resources of the Committee on Institutional Cooperation. The first report in April 1966 was made up of one-page abstracts of 30 instructional projects that were completed and judged to be successful by participating universities. The successive reports are now titled "Development and Experiment in College Teaching," and the 14 issues have followed essentially the same pattern.

These summaries are an operational definition of what teachers do to improve instruction. Although projects have dropped out, a rather significant number survive and provide valuable information about the characteristics of instructional projects that are likely to endure.

This series of reports is a unique contribution. Unfortunately, we've never had the funds to disseminate beyond the typical 3D list (deans, directors, and department chairpersons) on most of our campuses. The compendium is an appropriate model for future publications which will inform college teachers about significant instructional developments. Such reports must go beyond good intentions by including the conceptual rationale behind the procedure and evaluative information about the feasibility of the new arrangement and its impact on students.

Homework for Teachers

The essential educational moment comes about when a student sets herself to the task of learning something--usually from a book (broadly defined) and under the guidance of a teacher. This is the kernel of the matter, and the present chapter gives emphasis to information relevant to the teacher's influence on the student as the learner. We specify this purpose since, as mentioned earlier, so many published reports analyze and describe teaching as a free-standing process, or deal with issues and procedures relating to higher education only as a social institution. The intimate interactions among teacher, student, and a body of knowledge are brushed only briefly.

The kind of information important when preparing for teaching is brought into sharp profile when planning a course or a seminar on "college teaching." Formal or informal offerings of this type are sometimes used in large departments as part of a program for the orientation and preparation of the beginning teacher. (In this setting it becomes crystal clear that the first concern of the teacher, qua teacher, is "what shall I teach", and then, "how might this information best be presented to students?" "What shall I do as a teacher that is consistent with how students acquire information skills, and how they reorder their attitudes and values?") Information about instruction must be anchored to the student as the learner rather than to teaching as a self-defined process. This distinction between ends and means is the first consideration for planning and carrying out a formal program of information about teaching.

For three years I conducted an eight-week seminar for representatives from our larger departments who were involved with their TA (graduate student teaching assistant) program. As part of their homework, participants compiled a bibliography of what people in their discipline say to one another when they write about teaching. This was an interesting exercise which set forth the deficiencies and the strengths to be found when a mathematician, for example, writes for other mathematicians about college teaching. Similar patterns could be seen in the articles by professors of philosophy, English, biology, speech, etc. Fine discriminations about subject-matter organization and presentation were made along with global statements about the conditions for learning as experienced by the students. Nevertheless, having compiled such bibliographies, the participants felt that the informational bases for their departmental training programs for TA's were solidified.

Despite their persistent anxieties about how to teach, they were comfortable with this emphasis on "content" rather than on "method." Of the latter, they were suspicious and distrustful.

Given this content orientation, a university-wide "course" for prospective college teachers is less likely to succeed as a means for providing information. The level of discourse in such an interdisciplinary setting is abstract; teaching is presented in a general language. Substantive and procedural concepts are likely to be somewhat removed from the specific feelings of unrest experienced by the apprentice teacher. My own adaptation in working with beginning or experienced teachers is to emphasize factors under the control of the teacher that influence what and how well students learn. These are more pervasive and more powerful than the stylistic differences between teachers. Teachers appreciate gaining a clearer understanding about how verbal and other symbolic material takes on meaning for students, how they acquire concepts, and how information is stored and retrieved later in time and in quite different settings. These cognitive processes are basically the same from student to student in one subject to another. The generic relations between motivation and learning, between understanding and retention, for example, remain rather stable and do not change as a function of the course, the teacher, or the student's year in school. Here we have an appropriate informational base for teachers from different departments.

Teachers view their own task in their own terms; they make decisions about instruction which are aimed at being consistent with the content requirements of a particular course, the interests, abilities and background of information of their students and, inevitably, with their own individual personalities. Teaching is situation specific, and this poses a problem when writing or talking about teaching in the general case. The best resolution is simply to take advantage of the fact that college teachers are unusually intelligent and discriminating readers; they "process" information well and are quite capable of adapting an idea for application to their own classrooms. It is the responsibility of the information-giver to present ideas to these teachers which are worth their while--which sooner or later make a difference in what students carry away.

At the start, teachers may be looking for rather specific, concrete information--but frequently they insist on reserving their freedom to modify the recipe, to add the personal touch. A written description of a particular instructional arrangement will be more successful if the teacher has some understanding of the concept, the logic, and the data supporting the procedure being described. A report about teaching, therefore, should be more than a pedagogical show-and-tell. The teacher is the cook in charge of the kitchen, he is busy and expert in preparing an intellectual menu for students. Even so, a good cook also wants to know about the principles of nutrition. Understanding underlying concepts gives the teacher considerably more freedom and force in planning and providing food for thought for the students.

Jargon is often complained about. Nobody feels comfortable when pressed to use someone else's language for thinking about his own teaching. Systems engineers have some valuable things to say about teaching, and so do psychologists, sociologists, media specialists, psychiatrists, professional educators, and poets. There is no single party line for the analysis of good teaching.

But!

Reports about teaching will say more to specialists from different fields if written in a language which includes the student. We need to develop a discriminating set of terms, concepts, and procedures for analyzing the teaching-learning process --not a special pedagogical jargon, but clear distinctions, for example, between achievement and aptitude, reading and listening, learning and problem-solving, memory and imagination, thinking and feeling, motivation and emotion, reward and punishment, and evaluation and grading. Rather than giving lip service to what is best for students, teachers can actually bring to bear conditions which improve how students learn, remember, and reorder their attitudes and values.

A final word about physical things. First impressions make a difference, and this includes the morning mail. Most of these pieces, including junk mail, display the professional touch--printed, well-designed, and written with the crisp force of a TV commercial. The message may not be important, but it is well-packaged. The reports we receive about research or administrative matters from various offices in the university are usually well-presented. In contrast, information sent out about teaching is often mimeographed or dittoed in overruns beyond the limits of stencil sensitivity. If nothing else, these physical features remind us that pedagogy is a shoestring operation. And to this I object. Teaching is the central mission of the university, and information about teaching should be presented with the dignity and style it deserves.

There is, however, no point in a face-lifting operation if the message is sloppy, fuzzy, or picayune--or if it is self serving. We must remember that a university faculty is a demanding audience. To get their attention is one thing, but to hold it is more difficult. A report on teaching should end as soon as the point has been made--the only appropriate addition being the list of references to which the faculty might turn if they're interested in pursuing a topic in greater depth.

REFERENCES AND SUGGESTED READING

Axelrod, J. The university teacher as artist. San Francisco: Jossey-Bass, 1973.

Bass, R. K. & Lumsden, D. B. Instructional development: The state of the art. Columbus, OH: Collegiate Publishing, Inc., 1978.

Bergquist, W. H. & Phillips, S. R. A handbook for faculty development. Council for the Advancement of Small Colleges. Dansville, NY: Dansville Press, Inc., 1975.

Davis, J. R. Teaching strategies for the college classroom. Boulder, CO: Westview Press, Inc., 1976.

Davis, R. H., Alexander, L. T., & Yelon, S. T. Learning systems design. New York: McGraw Hill, 1974.

Eble, K. E. The craft of teaching. San Francisco: Jossey-Bass, 1976.

Ericksen, S. C. Motivation for learning: A guide for the teacher of the young adult. Ann Arbor, MI: The University of Michigan Press, 1974.

Ericksen, S. C. (Ed.). Development and experiment in college teaching. Committee on Institutional Cooperation--Panel on Research and Development of Instructional Resources. Ann Arbor, MI: Center for Research on Learning and Teaching, The University of Michigan, 1966-1979.

Ericksen, S. C. (Ed.). Criteria for the evaluation, support, and recognition of college teachers (Reports of the Fund Associates in National Project III). Ann Arbor, MI: Center for Research on Learning and Teaching, The University of Michigan, 1975-1979.

Gaff, J. G. Toward faculty renewal. San Francisco: Jossey-Bass, 1975.

Gaff, S. S., Feste, C., & Gaff, J. G. Professional development--A guide to resources. New Rochelle, NY: Change Magazine, 1978.

Guide to Effective Teaching. Report from the editors of Change. New Rochelle, NY: Change Magazine, 1978.

Kozma, R. B., Williams, G. W., & Belle, L. W. Instructional techniques in higher education. Englewood Cliffs, NJ: Educational Technology Publications, 1978.

McKeachie, W. J. Teaching tips: A guidebook for the beginning college teacher. Lexington, MA: D. C. Heath, 1978.

Pirsig, R. M. Zen and the art of motorcycle maintenance. New York: William Morrow, 1974.

Sheffield, E. F. (Ed.). Teaching in the universities: No one way. Montreal: McGill-Queen's University Press, 1974.

CHAPTER FOUR

SPECIAL FUNDS FOR THE IMPROVEMENT OF TEACHING

Robert H. Davis
Michigan State University

No matter how highly motivated faculty members may be to change instructional practices, they usually discover that personal commitment alone is not enough to bring about the change. Changes cost money. The amount may be no more than a few dollars, or it may be thousands of dollars, depending on what the faculty member proposes to do and how he proposes to do it. Faced with this problem, the typical faculty member begins to search for the needed funds, and sooner or later someone must decide whether or not the proposed changes are worthy of support.

All institutions fund a substantial part of their instructional improvement effort through regular administrative channels. A few institutions, having no special fund for the improvement of teaching, fund all instructional change through such channels. The faculty members in these institutions approach their chairman or dean for needed funds. Whether or not their requests receive a sympathetic hearing depends upon the priorities of the administrator, the availability of funds, and similar considerations.

An alternative practice is to establish a special fund for the improvement of instruction to make formal grants to faculty members. Numerous institutions have established special funds of this kind, and they are commonly judged to be successful. In his survey of 93 universities, John Centra found that 91 of his respondents used grants for developing new or different approaches to teaching, and 72 percent of those who said they had a program of grants felt it was either effective or very effective (Centra, 1976). A somewhat smaller percentage of the community colleges and four-year colleges surveyed had grant programs (78 percent and 95 percent respectively); 75 percent of community college respondents and 65 percent of four-year college respondents felt the program was effective or very effective.

Main Features of the Granting Mechanisms

Although many institutions around the country have developed granting mechanisms, no two are exactly alike. As a matter of fact, differences among granting mechanisms are as great as differences among institutions, because the design of the granting mechanism invariably reflects the history and organization of the institution in which it is embedded. The term "granting mechanism" is used in this context to mean a process for reviewing faculty proposals and deciding whether or not to fund them. In this article, I will describe how granting mechanisms differ and cite examples from a number of universities.

Five major considerations enter into the design of a granting mechanism: (1) centralization versus decentralization, (2) breadth versus depth, (3) institutional change versus individual recognition, (4) proposal evaluation by administrators versus proposal evaluation by teaching faculty, and (5) criteria for grants. Before discussing each of these issues, I would like to offer one caveat. While the four pairs of issues will be described as if they were categorical opposites, numerous options and mixtures lie between the extremes. Because it would be impossible to describe all of the various combinations and degrees of difference, I have taken the liberty of using this simple editorial device to highlight program design options.

Centralization versus Decentralization

All of the familiar arguments over centralization versus decentralization arise with the creation of grant mechanisms. At many institutions, the deans of the various colleges feel that money for instructional improvement should be allocated to their offices and that they should decide how the dollars are distributed. Two major arguments are commonly cited for decentralization. First, the deans have day-to-day contact with the instructional problems that exist in their colleges and are in a better position than a centralized agency to make decisions about instructional priorities. Second, centralization of such a critical activity undermines the authority of the deans.

The most obvious and compelling argument in favor of centralization is budgetary. There is simply no way that most universities can set aside significant amounts of money for every college to develop its own granting mechanism.

At the Big Ten institutions the total amount available for grants to faculty varies. The University of Michigan allocates approximately $85,000 per year; Purdue University, about $50,000; Indiana University, $100,000; Michigan State University, $110,000; and the University of Illinois, $70,000. One or two institutions in the Big Ten have set aside amounts substantially greater than $150,000, but the grants are not made solely for the purpose of improving instruction. The University of Iowa, for example, spends over $300,000 for semester developmental leaves and summer fellowships, but the leaves are for research, creative work, and

other scholarly activities as well as the direct improvement of instruction. Iowa also operates a special instructional support program, but it is much smaller.

A number of years ago, the University Senate at the University of Minnesota forwarded a resolution to the Board of Regents suggesting that the University set aside up to three percent of its instructional budget for grants for educational programs and instructional improvement. The size of the fund has varied considerably from year to year, averaging about $250,000 over the past eight years. This year the fund is about $200,000.

If these Big Ten institutions distributed grant funds among the colleges, the amount of money available to any given college would not be very great. Decentralized programs are less apt to achieve "critical mass" and the visibility with the central administration that some feel is necessary to guarantee their survival.

A second argument favoring centralization is that an all-university granting mechanism reflects an institutional commitment to the improvement of instruction. Sometimes an analogy is drawn to industrial research and development, the point being that every organization must set aside some money for the improvement of its most important products. If funds are distributed to colleges and departments, they are difficult to control, and teaching may not enjoy as high a priority in these decentralized units as travel, research, equipment, or other needs.

A third argument for centralization is that limited funds can be used most effectively as part of an integrated program that coordinates all available university resources to solve instructional problems. At a large university, it is simply not possible to divide instructional design expertise, media specialists, equipment, service, and so forth among the various colleges and still operate a first-class educational development program. The outcome of decentralization, so the argument goes, is likely to be underfunded, mediocre, piecemeal efforts to improve undergraduate education.

It would be misleading to assume that it is possible to assess an institution's commitment to the improvement of instruction by the size of the special fund it has set aside for this purpose. Although most Big Ten institutions have established special funds for the improvement of instruction, one or two follow other paths to this same objective. For example, although Northwestern University has no special fund for the improvement of instruction, it is nationally recognized for other approaches it has taken to improve teaching. It should also be noted that special instructional funds are only a small fraction of the large amounts of money universities earmark as their "instructional budget." The granting mechanisms being discussed in this chapter are resources in addition to the already existing budget support for teaching.

Even if an institution elects to establish a fund, the size of that fund can still be deceptive. Instructional improvement may be supported out of many budgets, including discretionary dollars available to administrators. Some institutions provide a wide range of media services without charge to departments, whereas at other institutions departments must pay for such services (See Chapter Five). Sabbatical leaves are often justified because they enhance the quality of instruction in the classroom. In other words, institutions may adopt very different strategies for the improvement of teaching, and the amount of money set aside for special grants to faculty is merely an indicator of the confidence an institution has in that particular option.

Breadth versus Depth

Given the limited number of dollars available for grants to faculty, should a program try to reach the maximum number of faculty possible, or should the number of grants be limited to as few as possible?

In practice, most institutions distribute available funds to a relatively large number of faculty in a wide range of departments. Purdue, for example, makes 15 to 25 awards per year. At Michigan State University the awards number from 20 to 35. Grants within the Big Ten range in size from less than $100 to $5,000 or more. In general, however, $5,000 is the upper limit; grants of this size are usually awarded to programs including faculty release time. Programs that stress depth, concentrating on a smaller number of grants, over breadth are more apt to make a significant and visible impact on a larger array of instructional problems than programs emphasizing breadth. The direct impact, however, will reach fewer departments, and this may decrease the chances that such programs will survive in a crunch. Distributing available dollars over many departments in different colleges naturally builds a large, broadly based constituency.

But a more substantive criticism of the in-depth approach concerns the generalization of innovations. Even when an instructional innovation is created for a single course within a department, it often fails to survive if the instructor who created it leaves the institution or has a change of assignment. Extensive curricular innovations are unlikely to generalize readily from one department to another, and those who attempt to diffuse them undoubtedly will encounter stiff opposition.

Institutional Change versus Individual Recognition

There are several ways to approach the granting of dollars to faculty members for the improvement of teaching. Many institutions pattern their procedures after federal agencies or foundations. Competitive bidding for available funds seems, at first glance, to be the wisest and most equitable policy; however, sometimes the results are disastrous.

Two years ago, I was asked to consult with an institution in Texas to help the staff establish an educational development program. Everything had gone well until the administration set up its granting mechanism. A committee of faculty had been established to review proposals that were submitted before a deadline. The committee had been asked to select and rank the best proposals. In the best foundation tradition, they established criteria and set about eliminating some proposals and accepting others.

By the time I arrived on the scene, serious doubts about the process had begun to surface. For one thing, some good faculty members were "turned off" by the process when their proposals were rejected. For another, some not-so-good teachers, who genuinely wanted to improve, were being told that their ideas had little merit, and so they simply retreated to their time-tested but unsatisfactory patterns of teaching. And finally, the members of the committee began to suspect that if they continued to operate in this way, they could expect increasing partisan pressure to get their departments a "fair share" of the available dollars. I was forced to conclude from this experience that granting dollars within an institution is not the same as granting dollars to an institution.

There is a deep philosophical difference between programs that award grants to individual faculty members based solely on the merit of their proposals, and programs that try to reach as many faculty as possible, regardless of the quality of their proposals. In the latter case, the objective is to develop the entire faculty and to move the whole institution to a new level of teaching. When a poor teacher comes for help, the objective is to help enlist his or her active participation in the change process. This is not to say that a program to improve teaching should fund any project, no matter how inadequate, but only that it should not "turn off" faculty simply because they do not rank first in the hierarchy. In other words, the goal is institutional change, not the creation of a teaching elite by selectively rewarding the best and most innovative ideas.

Obviously, there are problems with trying to fund as many proposals as possible. When questioned carefully, however, faculty members often discover that they can achieve their objectives with far less money than they originally thought necessary. Budgets must be scrutinized with great care to guarantee that every expenditure can be justified. Frequently the size of the budget for a given project can be adjusted to accord with its special promise or lack of merit. Finally, for a faculty member with a mediocre proposal, even a very small grant can "save face" and heighten interest in the improvement of teaching.

Proposal Evaluation by Administrators or by Regular Faculty?

Who should decide whether or not a project is worthy of institutional support? If the granting philosophy calls for supporting only those proposals most likely to produce a significant and lasting impact on instruction, the people best qualified to make this judgment are probably the teaching colleagues of the faculty member requesting the grant. It is not practical, however, in a

major university to assemble a different committee from every department to judge proposals. Even if it were feasible, it would be almost impossible to insure that criteria for grants were uniformly applied. Therefore, the most common approach is to appoint a committee of faculty from various departments for a limited tenure.

A small committee of more-or-less randomly selected faculty from several different departments may be no better qualified to judge a proposal than administrators, but presumably they have greater credibility than line administrators. On the other hand, they often lack the expertise in instructional design, evaluation, and media needed to make the soundest possible judgments about the merit of proposals submitted to them.

When the philosophy that underpins the granting mechanism emphasizes institutional change, administrators are able to play a more direct role in the process and overcome some of the disadvantages of the all-faculty committee. At Michigan State University, for example, every proposal is reviewed by a core committee of experts from the instructional development services. Often faculty members with special competence in a particular area are asked to review a proposal and sit with the committee. Sometimes department chairmen or deans are invited to attend review sessions. When proposals come before the committee, the faculty member is asked to participate in the discussion and hear the comments of those around the table. The faculty member may, if he or she chooses, bring a colleague or a graduate student to the review. When the time to make a decision arrives, the faculty member is asked to leave and told that the committee will shortly report its final decision. Ad hoc members of the committee are asked to stay and make their comments. Decisions are not made by vote; they are consensual.

I am often asked how such a system can possibly work. By the time the faculty member arrives at the meeting, he or she has discussed the proposal with technical experts in one of the instructional development supporting services; it has already been honed into the best possible shape. But more important, the group meets not so much to accept or reject this teacher's proposal (although some proposals are inevitably rejected), as to help the faculty member improve the quality of classroom instruction.

Criteria for Grants

There is considerable overlap in the criteria used to evaluate proposals. Some of the most common criteria used by the Big Ten institutions will be described in this section. Michigan State University has four basic criteria: (1) the number of students affected, (2) the experimental or innovative quality of the approach, (3) the generalizability of the approach to other academic areas in the University, and (4) the capability for evaluation. The University of Indiana's list includes some criteria that are not on the Michigan State list. Will the products or processes developed in the proposal have a reasonable chance to be used for a number of semesters, e.g., is the curriculum stable? Will faculty time or

other departmental resources be used more efficiently by implementing the proposal?

Most institutions use criteria to identify problems and analyze needs. At Purdue consideration is given to "the judged seriousness of the instructional problem with which the proposal deals, including, but not limited to, a consideration of the size of the student group that is likely to receive the proposed instruction."

The University of Michigan divides its projects into two groups and defines the criteria somewhat differently for each group. Faculty growth projects assist faculty members in acquiring new teaching skills or competence in evolving areas of their own or related fields. Instructional improvement projects emphasize the redesign of courses or curricula, or involve the development of new materials or learning opportunities for students.

Project continuation after funding ceases is a crucial issue for every granting mechanism, and most, if not all, programs require some assurance in this regard. Most institutions require that appropriate line administrators "sign off" at some point on proposals for several reasons: first, to draw to the administrators' attention the fact that a faculty member is undertaking a project that may require a long-range commitment of funds; second, to inform line administrators of the kinds of activities in which their faculty are involved. Faculty members sometimes appreciate the visibility that the administrative "sign off" gives them, and administrators become aware of the contribution of the special fund to their academic program.

Ohio State University requires that proposals be approved by a "learning consultant" as well as the department chairman and dean of the college. Another interesting caveat in its guidelines deals with copyrights and patents: "If a copyright and/or patent is to be registered as an outcome of the project, an agreement with the department and the Task Force must be arranged before the project is started." Michigan State University follows a similar practice; when grants are made, faculty are advised that products resulting from the effort fall under the policy, "Faculty and University Rights in University Sponsored Instructional Materials." This policy covers such issues as copyright, royalties, revision of materials, etc.

Although it is important for granting mechanisms to coordinate their activities with departmental, college, and university curriculum committees, this is frequently neglected. Grants often have a potential long-range impact on curricula. Since granting mechanisms are usually funded by the central administration, the very act of making a grant to a faculty member for a project with curricular implications may imply administrative endorsement of the concept. Later, if a proposal involving the grant is made to standing curriculum committees, the argument may be made that the university has supported this project with substantial amounts of money and, therefore, its approval should be relatively automatic. Needless to say, this argument seldom sits well with those charged with maintaining the quality of the curriculum. Therefore, when projects have obvious curricular implications, faculty members

should be encouraged to discuss their proposals with the appropriate members of such committees.

Objectives of Granting Mechanisms

All of these granting mechanisms attempt to create a climate of experimentation and change in the area of instruction; they often draw on a wide range of talents and supporting facilities throughout the university to improve instruction and diffuse effective methods across campus.

A well-designed granting mechanism may result in any one or all of the outcomes described here. It can provide the impetus to use instructional expertise on campus in a coordinated way. The request for funds legitimizes the involvement of experts from appropriate areas, such as learning, evaluation, computer science, media, etc. Specialists from such fields often have constructive ideas for the development of the faculty member's initial concept. Those responsible for the granting mechanism can play the role of "honest broker" by introducing the faculty members to others in the university who can provide assistance, including colleagues who have tried a proposed change.

A granting mechanism can also serve as a vehicle for faculty development. Faculty development involves a range of activities performed to promote the individual growth and development of teaching personnel. By encouraging appropriate workshops and facilitating communication among faculty and specialists in instructional development, the integrated granting mechanism encourages personal growth and learning on the part of everyone involved. In some cases, grants may be used more directly to fund faculty participation in workshops, seminars, and other learning experiences (on or off-campus) that help faculty to learn more about specialized methods and techniques or clarify their attitudes toward teaching.

Granting mechanisms can shape change in departments and colleges. For whatever reason, faculty in some disciplines are reluctant to become involved with instructional change. In many cases, the concepts and techniques in this area are not familiar to faculty members, who may be uncomfortable with them. If criteria for making grants are flexibly applied, lower demands can be made of faculty in departments and colleges that are reluctant to experiment with new approaches to the learning/teaching process. By adjusting funding criteria, grants can be made to faculty who would not otherwise become involved.

Finally, granting mechanisms can achieve three additional outcomes. First, grants reward faculty members and recognize their efforts to improve teaching. The symbolic significance of a grant to improve teaching should never be underestimated. Second, grants are used to buy needed equipment, software, or faculty time--all directly improve instruction. And third, a granting mechanism signals the entire faculty that the institution values good teaching and is committed to improving teaching.

How Effective are Granting Mechanisms?

Several years ago the Educational Development Program (EDP) at Michigan State University surveyed the faculty to whom grants had been made over a ten year period in an effort to evaluate its granting mechanism (Davis, Abedor, & Witt, 1976). Two questions in that survey focused on the issue of project survival. Faculty were asked, "Is what you developed as a result of the EDP project still being used in your department?" Eighty-one percent of the faculty answered that the project was being continued. The second question was, "Are you still using it?" Seventy-four percent of the faculty responded that they were still using the innovation on which they had worked. Many EDP projects apparently survive and find their way into the regular teaching process. From responses to two other questions, we can infer that an overwhelming number of project directors believed their projects were worthwhile (about 92 percent) and about 89 percent of all project directors had attempted additional innovations after they completed their EDP projects.

From the perspective of participating faculty at Michigan State, the MSU granting mechanism appears to be successful. Other data shed light on this. Jabker and Halinski (1979) recently surveyed the faculty and interviewed department chairpersons at Illinois State University (ISU) in an effort to determine how instructional improvement efforts were regarded and whether or not such efforts were rewarded. As in the case of the MSU study, faculty reported a high degree of intrinsic, personal satisfaction with their projects. When it came to formal, extrinsic rewards, however, the picture was different. They concluded that "despite a major instructional commitment to and wide scale faculty participation in instructional projects of various kinds, there is virtually no reward for such efforts" (Jabker and Halinski, p. 325). Like Feinker (1979) before them, Jabker and Halinski also found that many of those surveyed felt there was a major discrepancy between what ought to be true in this area and actual practices.

Promotion, tenure, and salary increase decisions are only loosely coupled to efforts to improve teaching. In part, this is true because teaching is regarded as a "professional obligation" that is presumably its own reward. Thus, a grant to improve teaching at ISU is not perceived as representing solid evidence of a faculty member's contribution to teaching; grants are merely funds provided to facilitate what many feel faculty ought to be doing anyway. If this view prevails, one would expect faculty interest in obtaining grants to fall off after a few years, and this is exactly what Jabker and Halinski found. Applications for grants at ISU declined by almost 50 percent from 1972 to 1975-76.

If good teaching is taken for granted, if sincere efforts to improve instruction go unrecognized, if extrinsic rewards are distributed for research and publication while good instruction must be its own reward, granting mechanisms and faculty development programs are, indeed, unlikely to have a lasting impact on the quality of instruction. I believe that the outlook is brighter, and that we are slowly--ever so slowly--convincing our colleagues that

the recognized discrepancies between existing practice and what ought to be the case are wrong and must be changed.

REFERENCES

Centra, J. A. Faculty development practices in U. S. colleges and universities (PR-76-30). Princeton: Educational Testing Service, 1976.

Davis, R. H., Abedor, A. J., & Witt, P. W. Commitment to excellence: A case study of educational innovation. East Lansing, MI: Educational Development Program, 1976.

Fenker, R. M. The incentive structure of a university. Journal of Higher Education, 1978, 4.

Jabker, E. H., & Halinski, R. S. Instructional development and faculty rewards. Journal of Higher Education, 1978, 49, 316-328.

CHAPTER FIVE

INSTRUCTIONAL TECHNOLOGY

Gene Faris
Indiana University

I hear and I forget
I see and I remember
I do and I understand
Confucius

Many institutions of higher education in the United States have established programs designed to encourage the use of media in the teaching-learning process. Reports on the scope and relative effectiveness of these activities are numerous. Conclusions range from statements that extol the virtues and unquestioned essentiality of the media to observations like that of Louis Leroy, a 16th Century French scholar, who described the effects of new learning and inventions in his own time, "All is pell mell, confounded, nothing goes as it should."

Without question there is a wide diversity of opinion concerning the use of instructional technology in the teaching-learning process in higher education; however, it is doubtful that the case is as good or as bad as the extremists in either direction contend. Certainly there are instances in which Mark Hopkins on one end of the log and a student on the other would characterize the most efficient and effective learning environment.

By contrast, there are teaching-learning situations in which the motion picture or television or the computer can make significant contributions to specific learning situations. These technological tools assist the teacher and student in overcoming the limitations of the classroom and their own experience. Through the use of educational media, one is able to experience phenomena that might otherwise be too small, too slow, too fast, too complex, or too distant to perceive. The media provide a concrete basis for conceptual thinking and can motivate and stimulate students.

Evidence to support these claims can be found in each of the 14 issues of Development and Experiment in College Teaching (CIC, 1966-79). Change Magazine, in its series Reports on Teaching, also documents many case studies in which media have made contributions (a common-sense evaluation as opposed to formal research) to the teaching of undergraduates. With evidence such as the above, I assume that media, properly selected and utilized, can make contributions to teaching and learning at the college level. Thus, the remainder of the chapter will focus on some of the problems and opportunities that the higher education community should attend to if instructional technology is to be used in the future in the most efficient and effective ways.

Resistance to Change--A Persistent Problem

At the college and university level, the lecture has been, and undoubtedly will continue to be, the most widely used instructional approach. One factor that has perpetuated the wide use of the lecture format is that an overwhelming majority of faculty members are comfortable with this technique. The same cannot be said about technological strategies. There are those who view anything technological as a threat to the status quo and, consequently, resist its introduction. It is interesting to note, however, that Ericksen, in his content analysis of the first 12 issues of Development and Experiment in College Teaching, says "about one third of all the abstracts reference the use of technological aids--computers, visual, audio, and multimedia. It is clear that turning to technology is an early option when teachers seek to improve a course of study; a careful reading of the individual abstracts indicates that teachers introduce these aids less to increase 'productivity' than to give students access to sources of information that might not otherwise be available."

Encouraging as this bit of information is, there are critics of the new technology who contend that these devices and practices will dehumanize education, placing emphasis on the machine rather than on learning. To the contrary, technology can personalize the educational process by freeing the teacher from the traditional role of lecturer-evaluator to a more exciting, challenging role of guide, counselor, resource. Aided by the tools of technology, the professor can adapt instruction specifically to the needs of the individual rather than to those of the group. An abstract in the Spring 1977 issue of Development and Experiment in College Teaching from the University of Wisconsin--Eau Claire campus provides an excellent example of this point. Professors Janke and Foust describe an audio-tutorial system in a basic earth sciences course in which technology has enabled them to respond to the needs of individual students. They also report that the audio-tutorial approach improved student performance and reduced the per-credit cost of instruction by approximately 25 percent.

This is only one of several examples that could be mentioned in which professors used the technology to serve their ends. The ultimate objective is to have teachers capable of controlling

technology and using it to their advantage, rather than have technology dictate what transpires in the educational setting.

How does one prepare professors to accept and use the newer, more elaborate technologies? How does one neutralize the impression of many faculty members that technology is a real threat to their very existence--that technology may somehow replace them? How does one present technology to the academic community as a friend, as a tool that can increase teaching-learning effectiveness? The answers to these questions are difficult, but one thing is certain: Faced with the greater availability of instructional technologies, colleges and universities must provide training programs to insure that faculty will understand and appreciate the capabilities and limitations of these new technologies. It is unrealistic to expect individual faculty members to achieve this goal alone. Institutions must assume responsibility for the design of programs to achieve such a goal.

The urgency of the problem is demonstrated by the fact that students who are preparing now to be faculty of the future will be teaching in the year 2000 when technology will have made its presence felt in a significant way in most classrooms in the country. Not only will future teachers need to be equipped to handle technology but the doctors, businessmen, lawyers, etc. of the future will be faced with a working environment that will require an appreciation for and knowledge of the same technologies that will be employed in the classroom.

For example, physicians in the state of Indiana receive continuing education programs through various combinations of the facilities of the Medical Education Resources Center at the Indiana University School of Medicine, the Indiana Higher Education Telecommunications System, numerous Instructional Television Fixed Service installations, and two cable companies. These facilities enable the Indiana University School of Medicine to distribute several hours of programming daily to hospitals around the state of Indiana. In addition, in those locations where there is an interface with a cable company, confidential medical programming can be received in the houses of physicians who are connected to the cable system. Add to the above the possibility of utilizing satellites to distribute such programming nationwide or worldwide, and one begins to appreciate the impact that technology will have on all segments of our society. Faced with this reality, college and university faculty have an obligation to do the best job they can in preparing students to function in an increasingly technological world. To accomplish this goal, faculty must be knowledgeable about and comfortable with the technological innovations at hand.

In efforts to create a more favorable climate for faculty use of media in their teaching, a number of approaches have been used in the Big Ten institutions. These have included programs designed to introduce faculty and teaching assistants to a variety of techniques and concepts and have ranged from sessions on simple production and utilization techniques to long-range in-service programs focusing on

various instructional developmental models. Some programs have encompassed the full range of concerns associated with the design of instruction, including the incorporation of media where appropriate. Another section of this report deals in greater detail with faculty workshops and other in-service techniques, so it should suffice here to say that the extent of media utilization in higher education will be closely related to the strategies employed to help teachers become at ease with the media.

Designing Systems and Facilities to Support Teaching

A frequent problem facing media professionals is the design and implementation of delivery systems, that is, having media materials and equipment in the right place at the right time. For example, the motion picture has been available for classroom use for approximately 70 years; however, many college and university faculty make little use of this valuable resource. Locating, securing, and showing a film is regarded by many as an annoying task. In many instances, a departmental chairperson will view using film as irritating and expensive and will discourage it. The same can be said for other media. If instructional technology is to make its maximum contribution to education, institutions must make a concerted effort to eliminate barriers, including the problems associated with inadequate delivery systems.

An adequate delivery system implies that certain costs associated with availability will be covered. Should instructional materials and services be provided through a central university fund? Should the department be billed for such services and materials? Or should the professor pay for such items out of his or her pocket? (This does happen.) Approaches within the Big Ten institutions vary considerably on this issue. For example, at Northwestern each school within the University is responsible for its own media support, which is financed in whatever way the dean of the school can find.

At Michigan State and Indiana University--Bloomington, many media services are provided at no cost to the faculty users. These free services include items such as films used from the local film library and student operators to show the films. Both institutions attempt to recover the cost of providing most production services; Indiana University--Bloomington has several subcenters strategically located around the campus which provide limited production services, thus encouraging faculty to use locally-created materials in their instructional programs. Other Big Ten institutions tend to operate in ways similar to Michigan State and Indiana. Such approaches reduce a number of barriers to the use of media in the college classrooms, and their advantages more than offset the difficulties created by individuals who seek to exploit a free service.

More serious difficulties are reflected in the teacher or student complaints that go something like this: "I wish we had a decent facility for showing films, filmstrips, and slides." Institutions of higher education are faced with the serious problem of providing adequate facilities for media utilization.

Unfortunately, many of the classrooms on college and university campuses were constructed years ago with little more than a room and a blackboard to facilitate the classroom lecture. To renovate such classrooms and make them more suitable for recent innovations in teaching can be quite expensive. Faced with this dilemma, it is absolutely essential that as new facilities are constructed, serious attention be given to designs that can accommodate various media. All too frequently decisions on building plans are made by individuals with insufficient information concerning the teaching-learning requirements of professors and students. Winston Churchill was correct when he said, "First we shape our buildings, and then they shape us." We must change this to: "First we determine our teaching-learning requirements and then design facilities to fit those requirements."

Purdue University's new facility, the Center for Instructional Development in the A. A. Potter Engineering Center, follows this guiding principle. First, it was determined that faculty needed the resources and facilities to undertake systematic course development activities. With this need at the forefront, a facility was designed that promises to become a landmark in the history of the University's efforts to support the teaching-learning functions of the Purdue academic community.

Although faculty and students can go it alone in the use of certain media, there are many instances in which the maximum contribution can only be achieved with the support of specialists. Few faculty members have the combined interest and expertise in subject matter, media development, and learning theory that the design of high-quality instructional materials requires. It is in this realm that most of the Big Ten institutions excel. These universities have established instructional development units which can provide the kinds of expertise necessary to undertake the systematic design of instructional materials, courses, or curricula. Michigan State University was one of the first institutions to establish a formal program to support educational development activities. Presently the Instructional Development and Telecommunications Services provides a variety of services to support the teaching mission of the institution. These services include:

1. Learning and Evaluation Service. Advises faculty regarding psychological, technological, or evaluation aspects of learning.

2. Instructional Media Center. Supplies faculty with free audio-visual equipment, delivery service, a film library, tape duplication, and graphics production at cost, as well as marketing University-sponsored materials.

3. Instructional Television. Provides free design, production, and distribution of credit courses for on-campus students.

4. Film and Multi-Media Production. Assists faculty in designing and producing educational, documentary, and public information films.

5. WKAR Radio. Broadcasts educational and service programs.

6. WKAR TV. Broadcasts educational and public service programs, including off-campus courses for credit.

7. Educational Development Program. Sponsors faculty-initiated projects and studies which will improve the quality of teaching and learning at Michigan State.

The Office of Instructional Resources at the University of Illinois at Urbana-Champaign represents another organizational arrangement for providing services that facilitate instruction. The staff provides assistance in numerous instructional areas including consultation on classroom instruction and management; construction, scoring, and analysis of achievement, placement, and proficiency tests; course, instructor, and program evaluation; research on problems related to instruction; and the design, selection, development, and use of media and other technological aids for instruction. For administrative purposes, the office is organized into these major divisions: (1) Course Development, (2) Measurement and Research, (3) Instructional Media, and (4) Instructional Television.

Other Big Ten schools have more recently developed such programs, and most are recognizing that for an institution to make a serious commitment to maximizing the potential of instructional technology, it is crucial that a team of highly intelligent individuals with differentiated roles be available to support faculty and students.

A Look to the Future

The new media technologies have profoundly affected other segments of society; can they be harnessed to serve the academic community? Technological advances such as videocassettes, videodisks, cable TV systems, satellites, and computers are exciting tools. Consider, for example, the immense storage and retrieval capabilities of the computer, or the satellite's ability to overcome time and distance barriers, or the availability of the videodisk system which can store 54,000 frames on one side of a 12-inch disk with the added feature of random access to each frame. These resources are here and now.

The computer, for example, is being used in many institutional situations. Much has been written about the creative work being conducted at the University of Illinois with the PLATO system. Ohio State University has experimented extensively with the Coursewriter system. A brief look at the titles of abstracts in Development and Experiment in College Teaching brings forth such interesting titles as "Computer-Managed Interactive Television in Undergraduate and

Life-Long Education"; "Computer Graphics in Undergraduate Education"; "Computer Simulation of Attitude Changes in Complex Social Settings"; and "CAI in Beginning Ancient Greek."

Satellites also offer some exciting educational possibilities. Recently a dramatic communication took place at the Indiana University School of Dentistry when faculty members held a joint meeting with their faculty counterparts from dental schools at the University of California at San Francisco and the University of the Pacific via the Communications Technology Satellite. The two-hour program offered a unique opportunity for the discussion of mutual concerns. Through the network of the Indiana Higher Education Telecommunications System, Dental School faculty members at campuses in Evansville, South Bend, Fort Wayne, and Gary also participated in this transcontinental exchange of information and opinions. The important consideration is, Will such activities go the way of their forerunners, where great promise was held for new and exciting innovations, with the actual results being rather shallow? The response of the higher education community to this question will depend on a number of significant factors. Some of these have been mentioned above; other closely-related considerations emerge with the use of these more elaborate technologies.

Cost effectiveness is a major consideration. For example, how does one compare the cost of conventional instruction to that of computer-assisted instruction (CAI)? There are projections that traditional instructional costs have been increasing at an average compounded rate of about 8.2 percent (Cable Television, 1974). On the the other hand, there are estimates that the cost of technologies for CAI are decreasing by about 5 percent per year, coupled with a 10 percent improvement in performance. These are interesting and valuable data; however, there still exists an urgent need for good evaluative experiments involving the new technologies that are conducted within a cost-effectiveness framework. A recent article in The Chronicle of Higher Education (Magarrell, 1978) provides data of the type needed concerning two computer teaching systems. The study, conducted by ETS, evaluated various aspects of the PLATO and TICCIT systems. Although the results of this study are not entirely supportive of the systems, it does provide the type of data needed to assist in generating data-based answers to cost-effectiveness questions.

Perhaps a second consideration is the need for software technology to keep pace with hardware technology. Unless powerful programs and materials are developed which fully exploit the unique capabilities of technological advances, innovations in hardware may suffer the fate of instructional television in the early 60's and computer-assisted instruction more recently. In too many cases, schools made considerable investments in sophisticated systems, only to learn that appropriate materials and programs were unavailable. Faculty initiatives in software development must be encouraged through institutional support if our investments in hardware systems are to pay dividends.

Having said all of the above about media in higher education, there are many individuals who contend that it would be rather naive to expect any great changes in the status quo unless more attention is given to the typical reward structure in higher education. It seems safe to say that many faculty members have the desire and talent to engage in teaching innovations involving the use of technology but are discouraged from doing so because the reward system does not give credence to this type of activity. I agree that when promotion and tenure committees attempt to evaluate the innovative uses of technology in the classroom, they are much less comfortable than when judging research and publication. I believe, however, that the problem is not so much with the system as it is with the evidence presented within the system. Thus, a plea is made for a greater effort in gathering data concerning effective and efficient uses of technology in teaching; such an effort will surely reap dividends and gain the recognition that many of our colleagues so richly deserve.

REFERENCES

Adler, R. (Ed.). Television as a social force: New approaches to TV criticism. Palo Alto, CA: Aspen Institute, 1975.

Alexander, L. T., & Yelon, S. L. Instructional development agencies in higher education. East Lansing, MI: Michigan State University Continuing Education Service, 1972.

Association for Education Communications and Technology. College learning resources programs. Washington, DC: Author, 1977.

Brudner, H. J. The past, present and future of technology in higher education. The Journal of Technological Horizons in Education, 1977, 4(3), 14-17; 22-25.

Burris, R. W. The CIC study of the non-traditional student, university resource supports for non-traditional study (Report No. 1). Iowa City: Division of Extension and University Services, The University of Iowa, 1971.

Carnegie Commission on Higher Education. The fourth revolution. McGraw-Hill Book Company, 1972.

Centra, J. A. Strategies for improving college teaching. Washington, DC: American Association for Higher Education, 1972.

Committee on Institutional Cooperation, Panel on Research and Development of Instructional Resources. Development and experiment in college teaching (Reports #1-14). Ann Arbor, MI: The University of Michigan, Center for Research on Learning and Teaching, 1966-1979.

Conference Proceedings, The First Annual National Conference on Open Learning in Higher Education. Lincoln: State University of Nebraska and Great Plains National Instructional Television Library, 1974.

EDUCOM. Cable television and the university. Proceedings of the Conference, 1974, Princeton, NJ: Author.

Ericksen, S. C. Development and experiment in college teaching at the Big Ten universities: A content analysis of issues 1-8 (Report #9). Ann Arbor, MI: The University of Michigan, Center for Research on Learning and Teaching, 1973.

Ericksen, S. C. Restless teachers and institutional support for change. Educational Psychologist, 1977, 12(2), 207-214.

Lerien, R. (Ed.). The emerging technology: Instructional uses of the computer in higher education (Report for the Carnegie Commission). New York: McGraw-Hill, 1972.

Magarrell, J. Computerized education. The Chronicle of Higher Education, April 26, 1976. (Comments by Robert E. Morris, Vice-President, Control Data.)

Magarrell, J. Computer teaching systems: Little impact on achievement. The Chronicle of Higher Education, October 30, 1978, p. 5.

Magidson, E. M. One more time: CAI is not dehumanizing. Audio-Visual Instruction, 1978, 23(2).

Merrill, I. R. Criteria for planning the college and university learning resources programs. Washington, DC: Association for Education Communications and Technology, 1977.

Meyer, J. H. F. Media innovation and faculty change. Audio-Visual Instruction, 1978, 23(2), 44-45.

Nolan, J. D. Are lectures necessary? Improving College and University Teaching, 1974, 22, 249-252.

Report on Teaching #1-6. Change Magazine, 1976-1978.

Smith, S. G., & Sherwood, B. A. Educational uses of the PLATO computer system. Science, April 23, 1976, pp. 342-352.

Thornton, J. W., Jr., & Bronn, J. W. New media and college teaching. Washington, DC: National Education Association, 1968.

Treas, C. E. How american business schools use media. Audio-Visual Instruction, 1977, 22(6), 18-19.

CHAPTER SIX

SEMINARS, WORKSHOPS, AND CONSULTATION SERVICES FOR FACULTY

Russell W. Burris
University of Minnesota

In the selection and promotion of faculty members, the Big Ten universities emphasize scholarship and teaching ability. Faculty members spend years preparing for their role as scholars; few faculty members, however, have had special preparation for their roles as teachers. It is not surprising, then, that a large number of faculty members at these institutions seek help with matters of instruction. Seminars and workshops for faculty members are common events within all the Big Ten institutions. It is to these forms of university-provided resources that this chapter is addressed.

Many faculty members seek and obtain advice and assistance for instructional improvement from their colleagues on campus and from disciplinary colleagues elsewhere. Departments and colleges often sponsor special presentations relevant to instruction; in addition, many standing faculty committees at departmental, collegiate, and university levels concern themselves with instruction. All of these represent significant sources of information for and assistance to the faculty member. This chapter, however, has to do with the assistance provided by specific offices and agencies as a regular part of their institutional responsibilities.

At each of these Big Ten universities, units offer assistance to faculty in dealing with questions or problems of instructional design and development. These agencies provide advice and assistance in the use and application of different media; in the construction of tests and other instruments for evaluating teaching and for assessing learning; in the evaluation of students, teachers, and programs; in providing funds for special instructional and curricular changes; in the research and development of alternative instructional materials and approaches; and in the dissemination of the results of instructional improvement projects. While the titles and the organizational

arrangements for such agencies are different, the aid and assistance available to faculty are similar among the Big Ten institutions. The similarities of these service functions are discussed in this chapter.

Librarians have been available to consult with faculty on instructional improvement for a long time, and some institutions established agencies to coordinate instructional uses of film as early as the late 1920's. In many cases the libraries and film-resource units established the precedent for assisting faculty in dealing with change and the improvement of teaching. As newer techniques and technologies appropriate for teaching and learning were introduced, expanded assistance for faculty was required. By the late 1950's, agencies and units such as those represented by the authors of this book began to appear on many campuses. These are the Office of Instructional Resources at the University of Illinois, the Office of Learning Resources at Indiana University, the Lindquist Center for Measurement at the State University of Iowa, the Center for Research on Learning and Teaching at the University of Michigan, the Development and Telecommunications Services at Michigan State University, the Center for Educational Development and the Consulting Group on Instructional Design at the University of Minnesota, the Associate Provost's Office for Instruction at Ohio State University, and the Measurement and Research Center at Purdue University. Although these units and agencies are not the only sources of assistance available--a few others exist at all-university, collegiate, and departmental levels--they do have as their major responsibility research, development, and coordination of instructional improvement on their campuses. Representatives of these units, as members of the CIC Panel, serve as sources of information for this report.

It is impossible in this chapter to describe all the ways in which Big Ten faculty members are assisted in their efforts to improve teaching and learning in their courses. Nor is it my intent to describe all the forms of consultation services, seminars, and workshops available to faculty members. Rather, my purpose is to summarize the consulting services at these institutions in order to identify and summarize those characteristics which appear to support teachers' efforts to improve teaching. The descriptions and interpretations are based on information provided by my fellow CIC Panel members and on my experiences in providing a consulting service for faculty and in directing seminars and workshops at the University of Minnesota over the past 14 years. My objective is to describe the ways, based on our collective experiences, in which the resources of consultation services, seminars, and workshops for faculty provide effective support for teaching and learning.

In terms of purpose and organization, this chapter asks the following questions: What are the problems of teaching for which faculty members seek assistance? What topics or issues are addressed through individual and group consultations, seminars, and workshops? What units and agencies provide aid and assistance? What are the different forms and arrangements for providing this support to faculty? How many faculty members seek such aid? What relationship does this form of support have to other institutional supports for

instructional improvement? What evidence of effectiveness exists for consultation services, seminars, and workshops for faculty?

Problems and Questions of Faculty Members

The teacher in the 1970's is presented with a number of challenges for changing and improving instruction not faced by teachers in earlier decades. The introduction of television, computers, and other technological developments has added complexities to instructional design decisions. Greater complexities have also come from the rapid expansion of knowledge in the disciplines. Students are required to learn more and to achieve higher levels of competency. Teachers are faced with difficult decisions about what is to be learned and what arrangements of courses can be used to handle the expanded knowledge. The explosion of knowledge in the disciplines has forced teachers to search for greater effectiveness and efficiency in teaching and learning. Not surprisingly, teachers ask whether the educational technologies could be useful in solving these difficulties, and if so, how.

Related to such developments, or perhaps because of them, is an increased research focus on the processes of teaching and learning. This research effort has identified not only a large number of variables which contribute to instructional effectiveness but also how the important interactions among these variables relate to instructional effectiveness (Burris, 1973; Cronbach, 1975; Jenkins, 1977; McKeachie, 1974). For example, whether the student reaches a particular level of achievement following instruction depends on what subject matter and skills were to be learned; what form, organization, and sequence were used to present the subject matter; what instructions, activities, and feedback were given to the student; and what abilities, interests, and prior knowledge the student brought to the learning task. Over the past 20 years the research literature on such instructional variables has grown at a rate comparable to that in most sciences and is a valued source of information. Phrased in one way or another, "learning theory applied to teaching" is a frequent topic in faculty seminars and workshops. The "laws" of learning cannot be sufficiently generalized to guide instructional improvement efforts independent of factors relating to the subject-matter area, student characteristics, and the achievement measures for particular courses. Hence consultation, workshops, and seminars are a helpful means for moving from the general to the situation-specific characteristics of a given teacher and course of study.

Several chapters in this book treat topics which frequently occur in working sessions with teachers: faculty development, student ratings, testing and grading, and the use of technological aids. The participants may have questions about specific, technical aspects of a procedure, but they nearly always appreciate the opportunity to examine the logic, theory, and research findings which support a particular line of instructional development.

Forms and Arrangements of Support to Faculty

As increasing numbers of faculty members seek aid and advice for improving their courses, the number of common questions and problems lead many of the units to establish seminars and workshops as an efficient approach in providing consultation services. In my own case, as I know it has been with others who have established seminars and workshops at Minnesota and at other Big Ten campuses, more reasons were involved for this formalization than the efficiencies gained through group consultation.

Many questions and problems brought to my seminar and workshop offerings by the Consulting Group on Instructional Design at Minnesota represent the surface of more extended and deeper issues. A department may, for example, ask for advice about enriching the laboratory experience of its students through film and/or computer-based simulations, or a faculty member may ask about alternative achievement measures to evaluate student performance in a course. To address these and many other similar questions related to a systematic instructional improvement effort requires the consideration of several other questions and issues.

Easy and straightforward answers to many of the faculty members' questions and problems are not immediately at hand. Any instructional improvement project aimed at being more than a temporary face-lift often demands involvement over an extended period. Many projects, for example, may require three or four years considering such questions as: What knowledge and skills from the subject matter are to be learned? What form, organization, and sequence are to be used to present the subject matter? What abilities, interest, and prior knowledge do students have that influence learning and achievement? Awareness of a teacher's need to deal with difficult instructional problems is a valuable concomitant of the seminar and workshop format. Indeed, the seminar or workshop atmosphere encourages analysis of complex problems and issues in college-level teaching.

One-on-one consultation is also provided both independent of and in association with seminars and workshops. Nevertheless, the trend over the past ten years is toward an increasing number of seminars and workshops.

Topics and Issues

At most of the campuses during the past several years, as many as 100 or more seminars, workshops, or similar forms of assistance have been annually available to faculty members. The number of topics represented in these sessions is almost as large. My own experience suggests that these seminars and workshops were established because many faculty members had similar problems and questions; the sessions were developed as a formalized consultation service. The following framework describes the major topics covered in these seminars and workshops: goals, objectives, and performance criteria; teaching

and learning activities; presenting the material to be learned; and students' characteristics.

Goals, Objectives, and Performance Criteria

The largest number of seminar and workshop titles reflect the priority teachers give to what it is that a student needs to learn and how to establish whether or not the student has learned it. Clearly, it is these issues of teaching and learning that represent the most fundamental challenges to the teacher. Effective instruction must reflect the teacher's ability at some level to describe and assess those features of student performance that distinguish "knowing" from "not knowing." Among the topics covered in many seminars are "establishing course goals and writing objectives," "testing and grading," "constructing pre- and post-tests," "computer-based tests," "criterion-referenced and norm-referenced evaluation," and "analysis for competence/expertness."

Certainly the issues raised by such topics represent basic complexities of teaching and learning. Indeed these issues are the focus of much of the research effort in current instructional science. Performance criteria, derived from clearly formed goals and objectives for particular courses, are necessary for the learner. Further, the degree to which course improvement is successful depends, in a fundamental way, on how effectively a teacher has dealt with these issues. The teacher of art history may want to teach students how to "look" at art and how to "see" similarities and differences among works in relation to different periods and cultures. In many ways, learning such basic skills of analysis and synthesis in art history is similar to the medical student's learning basic skills of visually analyzing and synthesizing a radiograph in relation to previous knowledge of physiology and pathology. While the goals and objectives of courses in these two subject matter areas have apparent similarities, the tasks and performances required of students at the completion of courses are different. The goals and objectives of every course must reflect what the student needs to learn in the course. In addition, the criteria used for evaluating tests and other measures of student performance must describe the degree to which the student has learned.

Presentation of Materials to Be Learned

Perhaps the largest number of seminars and workshops offered concern new and developing technologies for the presentation of subject matter. Often these sessions provide the faculty member with training and experience in the use of media for instruction, e.g., overhead transparencies, slides, audio recording, television, and computers. A few seminars can be found on each campus with a more specialized focus, e.g., film and slides in science laboratory instruction, television in second language learning, computer

simulations in medical education, and computer-based tests for evaluating clinical skills in legal education.

Usually these seminars or workshops are provided by specialists in the technology. In some cases, however, the seminar may be sponsored by units that stress broader issues of instructional design, such as research and development centers for instruction and testing, or they may feature teachers experienced in the use and application of technology for particular instructional problems, such as professors of law or medicine presenting instructional uses of video on the computer.

Teaching and Learning Activities

Many titles of seminars and workshops indicate that major attention is given to the arrangements of classroom and extra-classroom activities. A number of seminars include as topics lecturing, discussion, student-teacher interactions, dynamics of the classroom, and individualized instruction. In addition to these seminars, several institutions offer a micro-teaching service for analysis and critique, such as an opportunity to video-tape the teacher in a regular or simulated classroom situation.

In my own seminars, as well as those at other campuses, an attempt is made to relate the student-learning activities to instructional problems, e.g., the use of drills and practice in learning new vocabularies or concepts, the use of simulations in learning problem-solving skills, and the use of tests in providing feedback to the learner. Certainly, the design of teaching and learning activities is not independent of the subject matter to be learned, the nature of the student performance to be evaluated, the media selected for presenting the material to be learned, or the background and expectations of the student. Such problems and questions cause the faculty member to seek advice in improving instruction.

Student Characteristics

A few campuses offer seminars on the student characteristics related to course performance and achievement. The primary purpose of some of these seminars is the coordination of research on student development in particular educational environments, intellectual maturation of students, and learning style differences among students. They are concerned with how students' expectations, attitudes, and values relate to course performance and achievement and what factors influence changes in these subjective and personal states. Other offerings present local and more general research results describing various student populations and measures of student characteristics that influence learning.

Special Topics

On every campus special attention is directed toward the preparation of beginning instructors and beginning graduate student teaching assistants. At Indiana and Michigan, coordination and direction of this effort is provided by an all-university agency. At other institutions these units usually make presentations, but the coordination and direction is done at the departmental or collegiate level. Indiana University and Michigan State University provide all-university seminars for supervisors of graduate teaching assistants.

Other topics of a specialized nature among the seminar offerings at these institutions deal with the implications of the new copyright law for teaching, evaluations of instruction in promotion and tenure deliberations, and grant proposal writing.

Evidence of Impact

It is impossible to obtain an accurate estimate of the number of faculty members who participate in the seminars available on the various campuses. Based on the faculty participation data gathered from CIC Panel members for those seminars with which their units have direct involvement, the numbers vary between several hundred to about 2,000 annually on each campus. For example, at Michigan State University 1493 participated in the seminars and workshops offered by the Learning and Evaluation Service; this amount of participation represents a total of 5571 faculty contact hours. In addition, there are many other seminars and workshops offered by departments and colleges that are not included in these figures. In a single seminar offered three times a year, which I direct at Minnesota, approximately 100 faculty members participate for a total of 2,000 faculty contact hours. Participation in other seminars and workshops at Minnesota is at a degree similar to that at Michigan State, and similar patterns of participation are observed at Indiana, Michigan, and Ohio State.

With the exception of the few seminars offered especially for the beginning teacher, participation in these seminars and workshops involves assistant, associate, and full professors in equal proportion. Evaluation data are available for many of these seminars, and these indicate an almost unanimous endorsement from the faculty participants. A further indication of positive evaluation is the fact that the numbers of seminars and the numbers of participants have grown on each of the campuses over the past several years. It is also noteworthy that positive endorsement comes almost equally from all ranks.

Data concerning the costs associated with providing the seminars, workshops, and consultation services were difficult to obtain. Seminars and workshops are often only part of a unit's responsibilities, and in many cases much of what is presented in the seminars is volunteer effort. Nevertheless, on those campuses where estimates of cost were possible, it appears that between $75,000 and $100,000 per year goes toward providing this support of faculty. In

terms of faculty contact hours for these cases, the cost varied between $4.00 and $12.50 per hour.

Summary and Conclusion

Each of the Big Ten institutions providing information for this report shows evidence of support for faculty instruction with consultation services, seminars, and workshops. This support has grown over the past decade to a level representing a major commitment; evidence of its value clearly exists. The varied pattern of operation and administration of such programs among these institutions, however, suggests that there is a continuing search for effective ways to aid faculty members in their efforts to sustain and improve the quality of instruction. Given the nature of our institutions and the ways that different colleges, departments, and individual faculty members function, it is not reasonable to seek a single approach to instructional improvement. But it may be time to examine our collective experiences with seminars and workshops for faculty as important parts of a total instructional improvement program.

REFERENCES

Bowman, R. Observation and evaluation seminar, May 23-27, 1977. Evaluation Report submitted to Teaching Resources Center, Indiana University, June 16, 1977.

Burris, R. The "field plot" model for instructional research in higher education. Educational Psychologist, 1973, 10, 107-111.

Cronbach, L. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.

Ericksen, S. C. Restless teachers and institutional support for change. Educational Psychologist, 1977, 12 (2), 207-214.

Jenkins, J. Why it is hard to move from the laboratory to the classroom: A four-pointed problem. Paper presented at the American Psychological Association Annual Convention, San Francisco, September 1977. (Manuscript available from the Consulting Group on Instructional Design.)

Lasley, T. J. Report on task force on learning seminars for 1976-77: Seminar evaluations. Ohio State University, 1977. (Mimeograph)

McKeachie, W. J. The decline and fall of the laws of learning. Educational Researcher, 1974, 3, 7-11.

Michigan State University. Learning and evaluation service workshops. Michigan State University, Fall, 1977. (Mimeograph)

CHAPTER SEVEN

TESTING AND GRADING

Lowell Schoer
University of Iowa

There are few tasks which faculty enjoy less than testing and grading. The reasons are many and varied. At a philosophical level some faculty see little benefit from and the potential for great harm in the whole testing-grading process. More specific is the question of whether or not the return is worth the time and effort. Finally, one frequently grows uneasy when required to do a task about which there are many and often conflicting opinions and points of view; this uneasiness is likely to be intensified when the task is as important as testing and grading.

Obviously this chapter cannot answer all the questions or resolve all the conflicts that arise in testing and grading. An effort will be made, however, to clarify at least some of the issues of greatest concern. Although much of the discussion which follows is applicable to testing and grading in any educational setting, the focus will be on liberal arts colleges on large university campuses. The discussion will also be limited to testing and grading associated with assigning final course grades which become a part of the student's permanent record.

The information about specific grading practices on the various campuses was derived from a survey of catalogs of the Big Ten universities. Two approaches can be used to describe the results of such a survey. One is to summarize the information in some numerical form that indicates exactly how many institutions use each of a number of specific procedures. For this approach to be useful, it must be possible to identify and describe procedures with enough precision to establish categories into which individual institutional practices can be placed with a relatively high degree of reliability. Although the survey suggested that liberal arts colleges on the Big Ten campuses have similar philosophies of testing and grading, the actual procedures used are so diverse as to make it very difficult to develop such a set of categories. In

parts of this chapter which deal with specific practices, then, a second approach, which might be called the narrative approach, will generally be employed. This approach places much more emphasis on citing examples than on presenting numerical summaries.

Referent Used in the Grading System

A grade is a symbol used to identify a level of performance. As such, it has no meaning in and of itself, but it derives its meaning from the manner in which these levels are defined. Two general procedures can be used to define levels in a grading system. In one approach, levels are defined in terms of relative position in some group. An "A" might, for instance, be assigned to all students in the upper fifteen percent of a class, while a "D" would be assigned to all those in the lower ten percent. Such a grading system can be said to be "norm-referenced." The general descriptions of the grading systems and the terminology associated with the grades themselves indicate a very strong norm-referenced orientation in the basic grading systems used on all the Big Ten campuses. This is, of course, not surprising in view of the fact that this has been and continues to be the dominant approach to grading in all of higher education.

Although norm-referenced grades provide information useful for many purposes, they do not tell exactly what it is a person knows or can do. If grades are to provide such information, they must be associated not with levels that indicate relative standing, but with levels that specify competencies. This is referred to as the "criterion-referenced" approach to grading. The grades on a criterion-referenced scale are not defined in terms of rank in group, but in terms of performance standards.

Criterion-referenced grading has a number of advantages. It associates grades with descriptions of student behavior independent of group characteristics. In a given criterion-referenced system an "A" would mean the same thing whether it was assigned in an honors, an average, or a remedial section of a course and regardless of whether it came from a very selective institution or one that had an open enrollment policy. From the student's point of view, criterion-referenced grading has the advantages of identifying exactly what is required to attain a given grade and not making attainment of that grade dependent on the performance of others.

These are obviously important advantages. Unfortunately, criterion-referenced grading also presents some significant problems. Perhaps the most serious is that the nature of many of the outcomes we are trying to achieve in education do not lend themselves easily to the kind of performance-based definitions required for criterion-referenced grading. Although we often hear comments such as "That's an A-level paper" or "That's only C-level work," considerable evidence suggests that such categorizations are subject to a good deal of error.

Another problem associated with criterion-referenced grading is record keeping. For criterion-referenced grades to have meaning

on a student's record, those records must contain written descriptions of each category used. For a "B" in introductory psychology to be meaningful in a criterion-referenced sense, the record would have to include, for instance, a description of just what it was a student who received a "B" knew and/or was able to do. If we consider that such a description would be required for each level in a given course and multiply this by the number of courses a student takes, it is not difficult to see the magnitude of the problem.

Although criterion-referenced grading systems do present significant problems, the meaning they give grades more than compensates, especially where there is a heavy emphasis on the development of basic knowledge and skills or where the achievement graded is essential for success in some future activity.

Number of Levels in a Grading System

If a grade is defined as a symbol used to indicate a level of performance, there must be at least two grades in any grading system. Although the specific procedures and symbols used vary widely, all the Big Ten campuses make provisions for two-level grading under three different sets of conditions. They are (1) credit by examination, (2) "special" courses, and (3) student option in "regular" courses.

All the Big Ten schools make some provisions for credit by examination; without exception the grading of performance on these examinations is done using a two-level approach. A student does or does not receive credit depending on whether he or she scores above a previously established cut-off point on the examination. No grading distinctions are made either above or below the cut-off point.

A second provision is made for two-level grading when courses do not lend themselves easily to traditional grading or when the primary concern is to bring all students to some given level of competence. Independent study courses frequently fall into the first category. The second category consists largely of courses which involve the acquisition of knowledge and skills assumed to be basic to understanding and future success in an area; this second category represents by far the most common use of a criterion-referenced approach in the schools surveyed.

Courses requiring that students attain some specified level of competence raise the problem of what to do with those students who have not attained that level by the end of a grading period. On campuses such as Iowa, it is necessary to use a grade designed for other purposes, i.e., an "I." [illegible] though such use runs contrary to the stated purpose of the "I" ("To be used when a student's work during a session cannot be completed because of illness, accident, or other circumstances beyond the student's control"), no alternative is available.

A third provision made for two-level grading allows students a two-grade option in a course in which the regular grading system is ordinarily used. In two-level grading of this sort the levels in the regular system are collapsed and reduced to two. This requires, of course, that some decision be made about how this collapsing should be done, and there are some interesting differences in this regard among the Big Ten schools. Indiana and Illinois, for instance, both have what are essentially five-level grading systems. Under the two-level grading option at Indiana the top four levels (A, B, C, and D) carry credit. At Illinois the student must be in the top three levels (A, B, C) to receive credit for the courses taken under the two-grade option.

A number of advantages have been cited for two-level grading systems. Two of the most frequently mentioned are reducing student anxiety and encouraging students to enroll in course work outside their major area. A third advantage frequently cited is that a two-level system is a more "honest" reflection of our ability to differentiate among various levels of achievement in many areas. A fourth advantage, which is related to the assumption that we cannot differentiate between more than two levels of performance in many areas, is that the fewer levels there are in the grading system, the fewer errors will be made in placement. Finally, there is the issue of time. All other things being equal, the fewer the levels in a grading system, the less time it takes to assign grades.

With the exception of the possibility that a two-level grading system might encourage students to take course work outside their major area, these presumed advantages may be more apparent than real. While it is true a two-level grading system might reduce anxiety among students, the reduction may reach such a level that the student does little more than what s required to just barely get by. In addition, the anxiety students feel about grading may relate more to uncertainty about the criteria and the procedures used in assigning grades than to the number of levels in the grading system. Being explicit about these criteria and procedures may well do more to relieve pressure and threat than will juggling of the grading scale.

Academic achievement may be very difficult to measure; we should, however, recognize that this may be due more to our current level of skill in measuring than to something inherent in the nature of what we are measuring. If this is in fact the case, a two-level system should be viewed as a temporary solution to be used only until more precise approaches to measurement can be developed.

The reduction in errors in placement also has its price. Although decreasing the number of levels in a grading system reduces the number of misplacements, it makes each misplacement more serious. It is of little consequence if a student is assigned to the 73rd percentile when he or she should have been assigned to the 74th. It is a very serious matter, however, if the student is assigned a "Fail" when it should have been a "Pass."

The major disadvantage of a two-level grading system is that information is lost. All grading systems will involve some loss of information. There will be, for instance, a range of achievement among those students assigned B's in a traditional A, B, C, D, F system, and whatever information we might have about where a given student stands in a particular letter range is lost when we put him in that group. Loss of information may not always be a serious problem. Sometimes there are situations where the grade needs only to convey that a student has or has not (1) reached a certain level of proficiency or (2) been exposed to certain types of experiences; however, such situations are more likely to be the exception than the rule. In addition, we can never know exactly for what purposes grade information might be needed in the future. A user can always collapse grade levels if grading information is more precise than required for a given purpose, but it is impossible to recapture the information lost if collapsing occurred when grades were originally assigned.

This is not to say that there is no place for two-level grading in education. The nature of the achievement involved or the use to be made of the assigned grades might well make such a system a viable option. The option should not be exercised, however, without careful consideration of its cost.

If we want to go beyond a two-level system, how far should we go? Tradition has settled on five levels, and that is, in fact, the modal number used in the Big Ten. The basic grading systems at Purdue, Northwestern, Indiana, Iowa, and Minnesota are essentially five-level systems. Wisconsin has made a seven-level system by adding an AB between A and B, and a BC between B and C. Ohio State has an eleven-level system (A, A-, B+, B, B-, C+, C, C-, D+, D, E) and Michigan has thirteen levels (A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-, E). Michigan State has a nine-level system that uses numbers instead of letters (0, .5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0).

At a conceptual level, virtually all educational achievement could be assumed to lie on a continuum in terms of quality and/or speed of performance. Theoretically, then, there is no upper limit to the number of levels that could be identified in a grading system.

As a practical matter, however, there are two factors that must be considered when setting up a grading system. The first is the time it takes to do the grading. In general, the more levels in a grading system, the more time-consuming the grading process becomes, although for most courses going to a ten-level or twelve-level system demands no significant increase in faculty time. The second factor is the gain in the reliability of the information carried. In most cases, very little more is gained once the system goes beyond ten or twelve levels.

Increasing the Validity of Grades

Although the type of referent used, the number of levels established, and the nature of the symbol system employed are all important elements of a grading system, the meaningfulness of the grade assigned will ultimately depend on what information is gathered and how it is used in assigning grades. By far, the most common source of such information is some form of "testing."

As the term is used here, testing refers to any procedure designed to determine a student's level of achievement. It would encompass everything from the short daily quiz to the end-of-course examination, from filling in the blank in verbal statements to describing a course of action to follow in solving a complex problem in the laboratory. Although it need not be the only source of information used in assigning grades, testing is likely to be among the most useful.

Just how useful testing will be depends on how skillfully the faculty member develops testing procedures which are valid (measure what they are supposed to measure), reliable (give stable estimates of achievement level), and efficient (yield the maximum information about achievement level relative to faculty and student time required). Validity is generally assumed to be the most important criterion, but many testing decisions require trade-offs among validity, reliability, and efficiency. For instance, behavior that seems obviously to reflect a certain kind of achievement might well prove unstable or might require too much testing or observation time. In such a situation, validity may be compromised in order to gain reliability or efficiency.

The fact that testing frequently involves trade-offs among these three criteria could lead to the conclusion that the whole process is so complicated that any systematic test planning is virtually impossible. Actually, the complexity of the process only increases the importance of careful planning. Without such careful planning, testing cannot provide the kind of information required to make grades meaningful.

The first and most basic requirement for good test development is a clear idea of just exactly what the test is supposed to measure. Although there is a good deal of disagreement about how precisely educational outcomes should be expressed, there is general agreement that they should be expressed precisely enough to provide a basis for determining whether or not they have been achieved.

Secondly, to develop a good test one must choose testing procedures which are consistent with what the test is supposed to measure. In some cases there is an obvious one-to-one relationship between what the test is supposed to measure and a particular test item. If, for instance, a test is to measure the student's ability to organize and write a coherent paragraph, there is a very obvious and direct way of testing. Suppose, however, we want to determine whether or not a student can apply a certain set of principles in new settings. Then we must choose from a number of different item

types, any one of which could be used to measure this ability. To make the best choice, the faculty member writing the test must be aware of all of the alternatives, and the advantages and disadvantages of each.

Once the item type has been chosen, the item itself must be written. This might well be the most difficult and time-consuming part of the whole process of test development. In fact, no writing a faculty member will ever do is going to be read more carefully than are test items. It is essential, then, that the wording says exactly what it is meant to say, no more and no less. The difficulty of this task is compounded by the fact that items must be both precise and concise. If not, students may spend more time reading items than thinking about and answering them.

Finally, good test construction requires that the writer know something about item construction as such. What is the best form in which to present an essay item? How many alternatives should there be in a multiple choice item? How should the columns be arranged in a matching item? Where should the blanks be placed in a completion item, and how many should there be? Although questions such as these are largely technical, the faculty member's awareness of item construction techniques and his or her willingness to apply them will have a significant impact on the validity and reliability of the tests he or she constructs.

Improving Testing in the Grading Process

The improvement of testing practices on the campus requires that two things be done. The first is to create faculty awareness of the need for continuing effort to improve testing practices. The second is to provide faculty with support so that awareness can be converted into actions which will improve testing. One way that these can be done is by establishing an office or agency to which faculty can go for help in testing; all of the Big Ten universities have done this. Although the variability among the campuses makes it impossible to identity one approach as typical, that used at The University of Iowa can serve as an example of at least one way to provide faculty with testing support.

The Evaluation and Examination Service at The University of Iowa works with faculty on all aspects of test development, administration, scoring, analysis, and revision. Help in test development takes the form not only of workshops and short sessions on test construction but also of private consulting with faculty who request it. Once the test is developed, the Evaluation and Examination Service will type up and reproduce tests in whatever quantity the faculty member requests. If the test uses standard answer sheets, the service will score the test, provide item-analysis information, and accumulate scores for individual students across a semester. There is, in fact, very little that a faculty member needs to know about test construction or help that is not available from the service.

Faculty Support for Testing and Grading Systems

The improvement of testing and grading requires that we find ways to gain faculty support for and adherence to good testing and grading procedures. This is a difficult area because it touches on a number of sensitive issues related to faculty autonomy. The fact remains, however, that the grades which become a part of the student's permanent record are maintained and distributed by and in the name of the institution in which they were assigned. The manner in which they are assigned, then, must be an institutional concern. This does not mean that all grades must be assigned in the same way. It does mean, however, that grading practices of individual faculty members should be consistent with those established by the institution.

There are many reasons why faculty may not adhere to existing grading policies. In some cases, it is because they categorically reject the whole idea of one person judging another. In other cases, it is not so much a matter of philosophical opposition to making such judgements as it is a matter of compassion. It is not pleasant to have to assign students low or failing grades. In still other cases, faculty have found that it is possible to avoid a good deal of controversy by giving easy tests and skewing the grade distribution to the upper end. Fewer students are willing to argue that a test was too easy or their grade too high than are willing to argue that a test was too hard or their grade too low. There is also, of course, the very practical matter of keeping enrollments up, particularly when administrators are beginning to pay more and more attention to "semester hours generated." In a situation where the student is able to choose one course from among many alternatives, the assurance of easy tests and high grades might be just enough to tip the balance.

Each of these reasons relates more to the characteristics of the faculty member than to the testing-grading system itself, and no change in that system is likely to reduce their impact. In many cases, however, the problem is not so much faculty characteristics as it is that the grading policies themselves may be too restrictive. To say that an institution should have carefully defined grading policies does not mean that all courses must be tested and graded the same way; no one approach to testing and grading is equally appropriate for all the courses taught on a university campus. The diversity in the outcomes, content, and methodology of these courses precludes any such easy solution. Testing and grading policies must be flexible enough to allow for this diversity, yet should identify exactly what the options are and establish when and how they may be used.

Expanding the range of options available to faculty members should increase their compliance with the standards established within those options. The credibility of an institution's grading system is, however, sufficiently important that the institution should monitor testing and grading practices on a continuing basis. This does not mean someone continually looking over the shoulder of the faculty to be sure there is absolute compliance to a very rigid

set of procedures. The monitoring should, however, be done carefully enough to detect any continuing patterns of departure from the established standards. When such departures are found, the departments, programs, or faculty members involved should be given an opportunity to justify them. Exactly how this is to be done and what action should be taken if the justification is inadequate is something that will have to be answered at each institution, but procedures to deal with such noncompliance should be established. The institution, the faculty, and the students have too much at stake to make testing and grading practices a completely individual matter.

CHAPTER EIGHT

STUDENT EVALUATIONS OF INSTRUCTION

Warren F. Seibert
Purdue University

The Continuing Controversy

Among the issues in higher education, few appear as robust (virulent?) or as likely to achieve immortality as those associated with student ratings of instruction. It has now been a half century since the introduction of these rating methods (see, e.g., Guthrie, 1927; Remmers, 1928; Remmers, Shock, & Kelly, 1927). There have been hundreds of studies, reviews, and other entries into the literature of the field, yet most of its questions and issues can still light the fuse of controversy as quickly as ever.

Witness, for example, the article by Kossoff (1971-72) which contends that attempts to rate the quality of instruction and to develop meaningful statistics from ratings constitute a form of ersatz science. She approves of the objective improvement of teaching but believes that purely verbal description of teacher qualities is superior to any use of statistical summaries or profiles. Rather typically, Kossoff includes recollections of a former professor, one who would presumably be a poor candidate for favorable student ratings: "...a frowzy, middle aged woman with wispy drab-brown hair...narrow, piercing eyes...her voice came in sudden gasps...stern, humorless expression...unapproachable, both inside and outside class...she gave us...the finer nuances of Shakespearean scholarship that interested us not in the least." Then: "THIS TEACHER READS SHAKESPEARE MAGNIFICENTLY! ALL OTHER CRITERIA ARE IRRELEVANT!" (Emphasis in the original.) The Kossoff article drew prompt challenges from Eble (1972), also a Professor of English, and from Vattano (1972). In addition, it is worth noting that Kossoff makes no claim that classmates were similarly impressed; or, assuming that they were, she has not suggested a validity standard so high that cesium clocks and laser interferometers could be challenged with arguments of the same form. Precise though their measurements are, these instruments

are not free of error. Unquestionably, student rating results are fallible. They are subject to a degree of unreliability and, more seriously, to validity shortcomings, no matter how validity is defined.

Similar challenges can be directed toward practices that are nearly universal in academia and that have been largely immune to challenge. Undergraduate admissions criteria correlate rather modestly with students' undergraduate records, generally accounting for about one-third of the variance in students' later records. In graduate-professional school admissions, the accountable variance is still less. Thus, even the advanced testing technology of these times, based on decades of effort, provides procedures that can easily overlook the potential of a Darwin or a Verdi, both of whom were judged, in their time, to be marginal or unfit candidates. The absence of greater validity may be disturbing, but it does not follow that current practices should cease or be superseded by others of lower or unknown validity.

Controversy was sparked again when Rodin and Rodin reported a study (1972) bearing the subtitle, "Students rate most highly instructors from whom they learn least." The Rodins obtained a correlation of -.75 between calculus students' mean rating (on a single item) of eleven teaching assistants and the mean final examination scores of the assistants' students. The Rodins' report and a related later article by its senior author (1973) drew prompt response from Frey (1973, 1974), Gessner (1973), Centra (1973), and Menges (1973). Both Frey and Gessner include new findings which conflict--in Frey's case, quite dramatically--with the Rodins'. But in addition, the responding authors criticize the Rodins' work on these main points: (1) The teaching assistants were not the teachers of their course sections, but were assistants in fact as well as title and met with their students two days each week to answer questions and to administer or review test problems (Rodin and Rodin, 1972, p. 1165), while all 293 students received three lectures per week from the study's second author; (2) the final examination for the course was scored in a way which stressed detail over demonstrations of conceptual understanding; and (3) a competing interpretation of the Rodins' results is that students who needed and received the least help from the assistants "...would not only obtain the best grades but also rate the teaching assistants lowest in total teaching performance..." (Centra, 1973).* In addition, students were not randomly assigned to course sections, as good research design would require. While this appears not to have affected the Rodins' results greatly, its actual effects are unknown; this deficiency is especially noteworthy because it is

*Similarly, Remmers (1928) noted and commented on the fact that among instructors included in his study, some received ratings that consistently correlated positively with their students' "marks" across several traits or dimensions, while correlations for other instructors were just as consistently negative. This latter group seemed capable of "...flunking them and making them like it."

common in the literature, and it complicates the interpretation of many results.

There are still other indications of the controversy (e.g., Eble, 1972, pp. 56-58; Borgatta, 1970; Powell, 1978), but further demonstrations of this seem hardly necessary. Instead, it seems of more interest to consider the nature of the commonly raised objections. At least three broad classes of arguments seem to dominate. Kerlinger (1971) represents one class well, raising fundamental issues that concern faculty and student roles and educational purposes; Borgatta (1970) expresses similar concerns. Thus, if one shares Kerlinger's view that student evaluations undermine sensitive and critical relationships between faculty and students, "...are not an integral part of the instructional process...," and are not legitimate sources of motivation to improve, then it is doubtful that empirical work, however well executed and regardless of outcome, can influence these beliefs.

A second class of objections is based on the student rating/instructional evaluation literature, or portions of it, and interprets the evidence as showing that student ratings are seriously deficient in validity or other qualities generally considered important, even essential, for measurements to possess. Rodin (1971), Kossoff (1971-72), Gage (1961), and Powell (1978), all make some use of such arguments. In Rodin's case, it is ironic that she questions both the reliability and validity of ratings, since her own data, if not discounted, demonstrate the opposite. Her principal finding can properly be treated as a lower-bound estimate of the reliable variance associated with the single rating item used, since measures will rarely correlate more highly with external criteria than with themselves; the item yielded a high proportion of systematic--reliable--variance. Similarly, the validity represented in a coefficient of -.75, which Rodin reports, is the same as that of a +.75, so that the Rodins' results--again, if taken at face value--demonstrate considerable validity. To capitalize on this, however, one should take care to read the ratings, as it were, in a mirror.

A third type of argument appears rather frequently in oral discussions and may be more common there than in print. These arguments seem based on what Borgatta once termed "empirical thinking" (1954). This "methodology" seems to yield rationales and support for such conclusions as these: An instructor's popularity with students strongly influences or determines the level of ratings received; individual student's instructor ratings correlate highly with the grade these students receive or, in harsher terms, a quid pro quo relationship exists between individually awarded ratings and the student's grade; and the students' ratings lack reliability. The failure of research to lend significant support to such conclusions seems not to matter, and one can only speculate on the reasons. One possibility is that empirical thinkers find it relatively easy to imagine student "types" or even authentic cases whose behavior would seem to confirm such beliefs.

The continuing controversy is almost certainly fed by apparent inconsistencies in the outcomes of research and by the growing use of student ratings throughout higher education. While precise figures to reflect this growth are lacking, it now seems that a majority of American colleges and universities have either instituted such programs or are seriously weighing these decisions now. Bejar's national survey (1975) of 333 institutions shows 47 percent of the total sample and 68 percent of the respondents reporting the existence of a local student rating program; an additional 4-5 percent report that such programs are in the planning stages. His results seem to contrast with those of Astin and Lee (1967) and Seldin (1975), who in both cases surveyed only liberal arts institutions. A further indication of the prevalence of instructional rating programs is seen in Table 1. The authors of other chapters included here, each of whom represents a Big Ten university, have reported the status of student ratings at their institution. Their reports, as summarized in Table 1, suggest wide use of ratings even among large, research-oriented universities.

Institutional Decisions Concerning Use of Ratings

As institutions or their internal units consider the use of ratings, most discover that several related questions deserve to be considered. They come to realize that the deceptively simple question "Shall we or shall we not?" is coupled with many other, and frequently more important, questions. What, for example, are the alternatives to student ratings, and what practical or technical advantages do these have? Kulik and McKeachie (1975) briefly consider these: colleague ratings, administrator ratings, self-ratings, classroom visitation, and direct assessment of student learning. Eble (1972, pp. 59-63) considers similar alternatives, while Astin and Lee (1967) and Seldin (1975) include an expanded set, largely because of their broader interest in faculty performance in all roles, non-instructional as well as instructional.

The five alternatives considered by Kulik and McKeachie all have features that can recommend them, yet each also has practical or technical shortcomings which detract from utility. Compared to student ratings, their information is often difficult to acquire, and its reliability is generally lower, with visitation at a distinct disadvantage both because of lower reliability and the costs and practical problems of implementing such plans. Self-ratings also tend to be, understandably, lenient. In addition, although direct assessment of student learning has high credibility and may in time become more practical to implement, it remains cumbersome, and its information is rarely comparable across courses or among teachers. Thus, we would suggest that the several alternatives be seen as complementary, not competitive. Each can help to refine judgments of instructional quality, and none has features sufficient to recommend it alone. Furthermore, reliance on several information sources, with the kaleidoscopic patterns and contradictions that they will reveal, provides a valuable reminder that the arts of effective teaching are too numerous, varied, and subtle to be accounted for fully, much less simply.

Table 1

The Status of Student Instructional Rating Services at Seven Major Midwestern Universities

Institution	Student Rating System(s) Used *	Mandatory Use by Faculty?	For Diagnosis or Performance Review?	Extent of Use (approximate)	Additional Notes
Indiana University	Multi-Op and others	No, optional	Primarily diagnostic	Wide use of various instruments, originating within individual schools	The flexible Multi-Op system is relatively new and, at this point, has not been widely adopted by faculty.
Illinois, University of, Urbana-Champaign	ICES (Instructor & Course Evaluation System), CEQ (Course Evaluation Questionnaire), and others	No, but some form of student evaluation is encouraged	Both	2500 classes per semester; about half of all classes.	Ratings are regularly used in the publication of an "Incomplete List" of excellently ranked teachers.
Michigan State University	SIRS (Student Instructional Rating System), Levels 1,2, and 3	Yes, for SIRS levels 1 and 2; no, for SIRS level 3	Both	130-140,000 level 1 and 2 forms per semester; 4500 classes (?)	Faculty use of student evaluations is incorporated into University's code of teaching responsibility.
Michigan, University of	IDQ (Instructor Designed Questionnaire)	No, optional	Primarily diagnostic	1000 classes per semester	Faculty and faculty-student groups are undertaking to expand local uses of the relatively new IDQ services.

Institution	Student Rating System(s) Used	Mandatory Use by Faculty?	For Diagnosis or Performance Review?	Extent of Use (approximate)	Additional Notes
Northwestern University	Varies among schools of the University	Typically, no	Primarily diagnostic	Relatively low, but appears to be expanding	University Senate action asks all schools and colleges to initiate a student rating system; most units have now implemented their plans.
Ohio State University	SET (Student Evaluation of Teaching)	No, but some form of student evaluation is encouraged	Primarily diagnostic	Relatively low, but appears to be expanding	In 1975, University's Council on Academic Affairs approved implementation of a new, flexible SET system.
Purdue University	CAFETERIA	Typically, no, but in some units, strongly encouraged	Intenied to be diagnostic. Increasingly used in performance review.	1200-1700 classes per semester; 20-35% of all classes.	During six years of CAFITERIA use, faculty adoption grew steadily from the level of 300-400 classes to 4-5 times this level, even though most use was voluntary.

* The CEQ system at Illinois, Michigan State's SIRS, and Northwestern's several systems are essentially traditional; each employs a standard complement of items to comprise the rating instrument. Beginning with CAFETERIA, the other systems are flexible and computer-based (Multi-Op, ICES, IDQ, and SET) and they permit faculty to construct individually tailored instruments from an existing "pool" or catalog of items. Users may select 20 to 40 items, to which an additional few "core" items are normally added automatically.

If an institution is to incorporate student ratings in the instructional evaluation process or, as often happens, to encourage faculty in their use of ratings, what further questions deserve consideration? The more common questions include: (1) What are the likely consequences of mandatory and of voluntary faculty participation? (2) If mandatory, will evaluation be required in each course or only in some sampling of courses? A related question concerns "saturating" courses with ratings. If students are repeatedly asked to rate their courses and teachers, might they tend to treat the task perfunctorily? (3) If mandatory, is the institution prepared to defend this decision in the event of later challenge? Is it prepared to bear the burden of proof? (4) If mandatory, what precautions are needed to insure the quality or integrity of ratings? (5) Will results be provided only and confidentially to individual faculty or will administrators and perhaps others have direct access to them? (6) Will results be admissible as indications of individual teacher effectiveness and considered during promotion, salary, and tenure deliberations? (7) Are norms desired and, if so, what norm base(s) are to be used, or is it preferable to interpret ratings directly? (8) Are ratings intended (primarily?) for diagnostic use, as an aid to faculty in their instructional improvement efforts and, if so, what related services are needed (e.g., consultation, workshops, assigned time for course revision)?

Undoubtedly, there are other questions which can be important at individual institutions, but those listed seem sufficient to show the variety of choices to be made. What must also be recognized, even emphasized, is that few of the questions can be treated in isolation. Among the three principal uses of ratings--diagnosis-improvement, faculty promotion and related decisions, and students' decisions concerning courses or instructors--there are many conflicts or incompatibilities. (There is also a fourth use, research into the technical qualities and correlates of ratings, some results of which will be considered below.) The incompatibilities are such that one cannot expect a single rating system or service to satisfy each purpose equally. In fact, the opposite should be expected, since both the underlying motives of users and the desired technical qualities of ratings can change substantially as intended use changes.

Consider, for example, that while faculty may feel threatened when ratings are used in reviews of their performance or when these are reported in student publications, they feel little threat from the diagnostic use of ratings. Also, in the latter case, most participating faculty are motivated to discover shortcomings, but such results can prove damaging in the eyes of peers, administrators, or students. Moreover, the desired technical qualities of ratings tend to be different for each different use. For use in performance review, relatively little information is needed, since the principal objective is to refine judgments of general merit so that rewards or recognition may be allocated among candidates. However, if ratings are used by students in course-instructor selection, some descriptive detail is desirable so that

students can consider, for example, the relative merits of a highly organized course versus stimulating discussion. Some students may prefer one, while others prefer the other. Finally, for diagnostic use, faculty need reliable and specific information about modifiable aspects of their courses and instructional strategies.

In addition, norms are of variable value, depending on the use intended. If teaching performance is to be reviewed, most consider it helpful to know not only that a candidate's ratings are, for instance, high on an absolute scale, but that similar ratings are only received by perhaps five or ten percent of the faculty; however, for diagnostic use, faculty may well be satisfied to know the general level of each rating, without reference to others' ratings. Educators seem especially skeptical of student use of ratings in course or instructor selection (see, e.g., McKeachie, 1969), at least partly for reasons that Kerlinger (1971) discusses. There is concern, too, that ratings will receive excessive weight during reviews of faculty performance, largely because ratings yield numbers which can seem more precise than they are. Thus, ratings seem likely to be of most value when they are used voluntarily to diagnose instructional strengths and weaknesses or when they are one source among several used in reviews of faculty performance (see, e.g., National Project III 1977).

In view of the variety, complexity, and sensitivity of questions associated with the use of ratings, the importance of careful institutional planning can hardly be overemphasized. Institutions considering their use will surely benefit from investing the time and energy required to define local needs, to identify competing interests, and to decide which features constitute a service suited to local conditions, since decisions carry the potential for several unintended or secondary effects. For example, initially voluntary systems, if found to yield respectable results, can create pressures toward quasi-mandatory or even mandatory use; also, faculty who at first seek diagnostic information may decide that rewards are greater or more prompt when ratings are forwarded to the department head or promotions committee. Because such effects can substantially alter the role and worth of an evaluation service, the adopted policies and procedures must take such matters carefully into account. Needless to say, it is better to anticipate pressures and potential problems than to ride their currents unprepared.

Clearly, the position taken here is that ratings constitute a source of information which can be useful to individual faculty and to academic decision-makers. Emphatically, ratings are not advocated as infallible indices of good teaching. Rather, they are practical and systematic means for acquiring a set of reliable perspectives and judgments of instructional offerings. Clearly, too, institutional decisions concerning use of ratings are not to be made hastily or lightly. Institutional needs, like individual interests, vary, and to pretend otherwise is to run risks unnecessarily.

A Brief Look at Research on Ratings

The principal remaining question concerns the technical properties and correlates of ratings. What does the literature indicate and what confidence can users place in the results of ratings? It should be recognized at the outset that broad generalizations are difficult for several reasons. First, rating procedures and devices vary greatly, encompassing the single item of Rodin and Rodin, the extensively developed Course Evaluation Questionnaire (CEQ) from the University of Illinois, and others with similar features and origins, plus many of intermediate coverage and/or technical development. In addition, groups of studies that have sought to address common questions often employ different general methods and are executed in different surroundings or circumstances. Such differences can not be ignored, but neither can they be easily incorporated into generalizations that seek to accommodate mixed results. Also, many of the empirical questions that concern ratings are far easier to conceive and to design as studies than to execute faithfully. Often, the unit of analysis is comprised of sections of a given course and/or the instructors of these sections, thus requiring a high level of cooperation from an investigator's colleagues, as well as an inherent flexibility to meet the requirements of each study's design. In the face of these difficulties, compromises are often made, and ambiguity is introduced into study results.

In considering the principal findings from research, we shall paraphrase and abbreviate the seven rather broad questions considered by Aleamoni (1974b). In condensed form, his list of prevalent questions includes these:

1) Do students possess the maturity to make consistent judgments of instruction? Similarly, do students' judgments generally coincide with those from experienced faculty?

2) Are students unduly influenced by instructor popularity, especially popularity acquired by awarding undeservedly high grades?

3) Are ratings not only reliable but also valid indicators of effective teaching? Particularly, do ratings correlate positively with external measures of student group (or class) learning?

4) Are ratings substantially influenced by such extraneous factors as class size, course level, and whether the course is required or elected?

5) Do student ratings serve to aid faculty in the improvement of their teaching?

Several reviews that encompass major segments of the student rating literature have appeared within recent years, and these examine most, if not all, of the foregoing questions. Readers may

be particularly interested in the reviews of Remmers (1963); Costin, Greenough, and Menges (1971); Kulik and McKeachie (1975); and Doyle (1975); as well as Aleamoni. Also, readers opposed to ratings may enjoy Powell's article (1978), which examines a variety of studies and conclusions and seeks to explode the "...Rating Scale Validity...Myth." These authors and the studies to which they refer present most of the relevant findings. In addition, however, there are a few recent studies that merit consideration and that bear particularly on questions three and five. Thus, the plan here is to (1) consider five fairly broad questions, much as earlier reviewers have done, (2) to examine a sampling from the same collection of empirical studies, and then (3) to supplement this with references to some recent studies.

Question One--Student Judgment

The first question asks whether students have the maturity or experience to render consistent judgments of instruction and whether students' judgments also coincide with those of a faculty member's peers. Consistency can be and has been examined in various ways. Do students enrolled with individual instructors during successive weeks or semesters rate their instructors similarly across time? Do random halves or random student pairs from instructors' classes award highly similar (correlated) ratings? Do alumni ratings coincide with those from current students? Do instructor and/or course profiles of strengths and weaknesses show stability across different sections of an instructor's course?

The repeated outcome of many such studies supports the conclusion that most ratings are quite satisfactorily consistent or reliable; coefficients in the range of .75 to .85 or .90 are common (see, e.g., Guthrie, 1954; Spencer & Aleamoni, 1970; Seibert, 1977). Drucker and Remmers (1951) also report correlations of .40 to .68 between current students' ratings of instructors and those of alumni ten years removed from their undergraduate experience. To some extent, discrepancies between the groups' judgments must reflect changes occurring during ten intervening years, rather than simply the insensitivity of either group. It is also worth mentioning that reliability is sometimes confused with student group consensus (Rodin, 1973), as if a condition for reliability was that all students in a class should agree substantially in the ratings they assign, e.g., to their instructor. Yet it is possible for an item or instrument to yield reliable average ratings, say, from random halves of a class or across occasions, but still to elicit diverse ratings from individual students. "My instructor motivates me to do my best work" need not show strong consensus, although the average rating from the class and even the responses of individual students may be highly stable. In short, one should expect some instructor or course characteristics to interact with individual student characteristics, thus yielding diverse, yet potentially reliable, responses.

The second part of the consistency question asks whether students' judgments coincide--are consistent--with those of a

teacher's peers. Do faculty and students hold similar views concerning which faculty are the better and poorer teachers? The studies of Maslow and Zimmerman (1956) and of Hildebrand, Wilson, and Dienst (1971, p. 7) adequately represent those that address this question. Both show substantial agreement. In the first of these studies, the correlation between students' and peers' ratings of the same teachers was .69 and in the second study, faculty and students both nominated "best" and "worst" teachers. Of 66 teachers named by both groups, there was agreement on 56. As Kulik and McKeachie (1975) note, however, such agreement is not necessarily the result of independent judgments since "...grapevines stretch up as well as out..." and teaching reputations are part of the information grapevines carry. Note, too, that to the extent that faculty peers' and students' judgments agree, there are problems in challenging the worth of one without also challenging the other.

Question Two--Instructor Popularity

The next question arises from one of the most common objections to student ratings, which is that they reflect little more than instructor popularity and, worse yet, that students will award high ratings in return for easy grades. The evidence lends little support. First, the eleven factor analytic studies summarized by Kulik and McKeachie show that rating instruments typically "decompose" into several factors or rated dimensions of instruction, not one. Of four factors which appear with good consistency across different studies, one, which was labelled "rapport," has popularity overtones. Nevertheless, rapport is more readily equated with instructor empathy, approachability, or friendliness than with pandering or reciprocal desertion of principle. Three other factors were labelled "skill," "structure," and "overload," although in individual study reports, investigators chose such labels as "instructor competence," "organization," and "academic emphasis" to name the apparently similar factors. Thus, the evidence indicates that students discriminate along several dimensions of instruction, rather than only on the basis of instructor popularity, entertainment, or other dubious qualities.

The related question of whether students award high ratings in return for high grades, and vice versa, is examined by Costin, Greenough, and Menges (1971) and by Powell (1978), with different conclusions reached. In the first of these reviews, the authors cite about a dozen studies in each of two categories: those finding no relationship between students' ratings and their actual or expected course grades and a second group in which significantly positive correlations were found. The latter correlations are generally low and they "...might better be viewed as a partial function of the better achieving student's greater interest and motivation, rather than as a mere contamination of the validity of student ratings" (p. 519).

Powell (1978), on the other hand, is quite convinced that, metaphorically, the container is at least half empty. He scores points for having ranged across a considerable body of literature,

but loses credibility as he treats questions and data in ways that are, to say the least, novel. On page 620, for example, he states: "There is evidence that grade-rating correlations are likely to be highest in departments such as English..." but cites no evidence. Continuing, "A difference of a half rating level would certainly result in a difference of five deciles or more...on a typical rating scale." He seems to mean that an instructor whose average global rating is, say, 4.0 will be at least fifty percentile points (five deciles) above a colleague whose average is 3.5. Since he elsewhere refers to typical instruments as those using 5 or 6 point scales and since those used locally (and at a few dozen other institutions) qualify, we checked local norms* and found that a rating of 3.5 corresponds to the 27th percentile, while a 4.0 converts to the 54th percentile; the two hypothetical instructors differ by half as much as the minimum which Powel claims. At all other points on the score range, the percentile (or decile) differences are even less (from 2.5 to 3.0 corresponds to 8 percentile points; from 3.0 to 3.5, 13 points). Still, on page 620, he cites Hudelson's 1951 (or 1952) study, the date seeming to depend on which of Powell's citations one accepts. The reported rating-grade correlation is .19, but Powell then re-summarizes the data to show average ratings given by A, B, C, and D students, creating an appearance that the correlation is 1.0, since averages decline regularly from A through D. Nevertheless, an r of .19 still accounts for less than 4% of the variance, not 100%, as implied. Such tactics assure that any linear correlation, however small, will appear superficially to be perfect. Powell's interpretation further confuses matters, since he has just given data in which students were categorized by the grade they received, but he now groups them by their teachers (or intact classroom groups) and says, "The teacher who gave an average grade of 'B' might, <u>therefore</u>, [emphasis added] expect to rank many deciles higher than the teachers who gave an average grade of 'C'." Neither his logic nor known data support this assertion.

Powell then takes four lines to discuss yet another study (Bassin, 1974) which he regards as supportive of his views. He reports that Bassin has found an r of .10 between students' grades and their instructor ratings which "...symbolized a 32 percentile difference in rankings between teachers who gave higher grades and those who had given lower." Instead, Bassin has in fact reported an R^2 of .10, and the relevant related statistics, summarized in the first two rows of his Table 2, show that the predicted ratings of more and less "lenient" graders (teachers) differ by 9 or 15 percentile points, depending on whether one contrasts those who teach "quantitative" or "non-quantitative" courses, two terms that are employed without accompanying definition. Moreover, one must surmise that Bassin has executed a multiple regression analysis of a dubious kind, since there is no indication that R has been shrunken or cross-validated, n (which is 63) is minimal for a two-predictor analysis, and the five criterion variables are treated as if uncorrelated.

---

*The global item's contents are: "Overall, this instructor is among the best teachers I have known".

A few more comments, then the Powell article will be put aside. The strong implication is given, especially on page 620, that correlation is causation, i.e., that a relationship between grades given and ratings received is such that faculty can and do directly manipulate ratings through the grades they assign. This is a simplistic insult to students and faculty and, as a general proposition, it clearly is false. Moreover, in the literature one finds no persuasive evidence for such a view. Also, on page 625, Powell takes Frey (1974) to task for changing "the rules" and employing rating items unlike those found elsewhere. (Frey's items appeared to yield measures highly correlated with his measure of instructional effectiveness.) Frey's work can be questioned on other grounds, but not for honoring a tradition as old as psychological testing itself. Much work in this field has routinely taken its validity wherever it could be found; where sets of variables are usefully related, they are frequently incorporated into prediction, placement, diagnostic, and similar programs.

We therefore suggest that, in view of the findings from factor analytic studies, which repeatedly identify several dimensions of student judgments, not a dominant dimension recognizable as teacher popularity, one must be suspicious of claims that students judge instruction in simplified terms. In addition, the relationships found between the grades students receive and the instructor ratings they assign are generally modest. We suggest that such relationships are best understood as Costin, et al. have indicated, rather than as evidence of an unholy alliance between students and their teachers, each seeking immediate advantage through the short-circuiting of principles. It would be odd indeed if students who have done well academically did not, in general, have positive perceptions of those who have taught them.

Question Three--Ratings as Valid Indicators of Effectiveness

The third question is as important as it is difficult to resolve. Are ratings valid indicators of instructor or instructional effectiveness? Or, do students rate most highly the instructional experiences from which they gain the greatest educational benefits? A brief look at relevant literature will show that many difficulties complicate such questions, a few of which are these: (1) How is effectiveness or benefit to be defined--as students' examination scores, as skills retained over extended time periods, as motivation or interest in learning, as net increase in skill, or perhaps as student learning directly attributable to the teacher? (2) Even the best measures of learning or instructional effectiveness only approximate the extent of education's effects; ultimate criteria are lacking. (3) When validity is defined in customary terms, i.e., using some measure(s) of student group learning, the most defensible research design calls for random assignment of students to instructors/course sections, yet this procedure poses practical problems which are rarely surmounted. (4) When several instructors are responsible for sections of a single course, they often interpret course goals differently, with the result that no measure of effectiveness can be equally fair to every

instructor. (5) If, as some contend, most instruction is at least adequately done, then the range of effectiveness is restricted and, with this, the likelihood of a high correlation between instructional quality and any other variable is reduced. (6) The validity of ratings has been studied throughout a time period in which higher education and rating techniques have been changing (some would say "developing"), yet one cannot judge whether these changes have affected findings systematically. (7) Validity of ratings, as usually defined, may in fact vary systematically and as a function of teachers' academic status or experience, (which some data suggest, see Gage, 1961; Sullivan & Skanes, 1974), of content taught (see McKeachie, Lin, & Mann, 1971), or of other variables.

Such difficulties have not prevented the accumulation of a considerable body of findings, but they certainly complicate interpretation. The plan followed by most investigators is to identify one or more multiple section courses, then administer an end-of-course ratings instrument and a common examination in all sections of each course. Usually, examination scores from the sections are statistically adjusted to minimize effects of pre-existing group (section) differences. Then, adjusted scores are correlated with average ratings awarded by the students in the various sections. Procedural variants include use of grades in place of adjusted scores and, less frequently, random assignment of students to course sections (see Sullivan & Skanes, 1974; Centra, 1977).

From information already presented, readers can probably anticipate what validity studies have shown. First, as the Rodin-Frey-Gessner debate reveals, ratings of the general quality of instructors or courses sometimes correlate positively, sometimes negatively, with measures of student (group) learning. More often than not, the obtained correlations are positive but of modest size. Our own conclusion, which other proponents of ratings share, is that students will tend to rate most highly those instructors and courses from whom or which they learn most. This tentative conclusion is immune to the Rodins' results because of methodological and conceptual weaknesses described earlier. Nor is it affected by Frey's and Gessner's distinctly positive findings, since these studies are also flawed. More persuasive is the work of Sullivan and Skanes (1974), which correlated group (average) ratings of instructors with group (average) achievement and found an average correlation of .39 across 10 courses that represent 6 disciplines. Centra's 1977 study is similarly persuasive in outcomes and design. Centra reports correlations of 9 rating variables with group (average) achievement for 7 courses from 5 disciplines. His overall ratings of teaching and of the course's value yielded Pearson or partial correlations of -.32 to .96, with 8 out of 14 of these reaching .58 or higher. Correlations of other rating variables with measured achievement were generally lower; however, the median of 63 principal correlations reported (7 courses times 9 rated variables) is about .41.

These two studies are especially noteworthy because each describes a highly credible criterion measure, each incorporates data from several disciplines, both used random assignment of students to course sections (although Centra managed this in only two of seven courses), and both reflect recent conditions in higher education. Readers will of course judge these and similar studies as they wish, but our qualified conclusion is that a useful, if modest, level of validity is normally present, even if a demanding definition of validity is applied. Furthermore, justifications for the judicious use of ratings can derive from these considerations: (1) Alternatives to ratings provide no practical or technical advantages that establish their superiority. (2) Students' judgments of their educational experiences deserve some measure of respect on the grounds of face validity. (3) To put it not too bluntly, if it is acceptable to academics to apply student selection/admissions criteria which have only modest validity, to evaluate students' academic performance with measures not noted for their reliability (or empirically determined validity), and to evaluate faculty scholarship and/or creativity by means that have face validity but little else to recommend them, how much more can reasonably be asked of evaluations from students?

Question Four--Influence of Extraneous Variables

The fourth question asks whether ratings can be influenced by extraneous variables, i.e., those that are largely or totally beyond instructor control. Do class size, course level, student interest in subject matter, or other similar variables appear to affect students' ratings of instruction? In general, the answer is that they do, but the evidence includes further inconsistencies. Class size, for example, seems to have some slight influence (see, e.g., Heilman & Armentrout, 1936; Lovell & Haner, 1955), yet Guthrie (1954) found none. Moreover, Gage (1961) describes findings in which lower ratings are associated with classes of moderate size, rather than with larger classes, as other evidence suggests and as most would suppose. Gage also reports that faculty in the two higher academic ranks fared better than those in lower ranks. Some suggestive, but indirect evidence is also reported by Sullivan and Skanes, who found that the ratings of full-time psychology instructors correlated .53 with their students' measured achievement, whereas the comparable r for part-time graduate instructors was .01. In addition, r for "experienced" (one year or more) full-time instructors was .68, while for "inexperienced" instructors, r was .13. Average ratings for the contrasted groups are not reported, so one cannot conclude that experience or rank influenced rating levels, but these variables seemed related to the judgment process and seemed to affect the meaning attributable to results.

Elsewhere in the literature, investigators have examined a variety of other variables that instructors have little or no power to control--instructor personality, the discipline the instructor represents, sex of the rater, and students' pre-existing levels of interest or motivation. All such variables appear to have some

potential for influencing ratings. Typically, their influence is not great and, as the class size studies suggest, specific findings can prove difficult to replicate. In view of such results, there is further reason to interpret and use rating results sensitively. As with many things, part of the meaning derives from a context, and interpretations divorced from context may lead to oversimplification. Interpretations may also be improved through use of discriminated norms. Kansas State University's IDEA system provides an example of such norms; it permits each instructor to have his/her ratings compared with those of others who teach toward similar goals, with similarly motivated students, and in similarly sized classes. While discriminated norms do not remove the need to exercise care in drawing meaning from ratings, they can help to accommodate for systematic differences, where these are known or reasonably suspected.

Question Five--Ratings as Means toward Improvement

The fifth and last of the questions concerns the value of ratings as means toward instructional improvement. Interest in the question is relatively recent, the number of relevant studies is small, and their results are not conclusive. Of the known studies, Centra's (1973), McKeachie's (1975), and Pambookian's (1974) seem to show that knowledge of one's earlier ratings has little effect, perhaps none, on subsequent ratings. For Centra, there were no significant differences between the two principal instructor groups, although there were some small improvements noted for those instructors who expected higher mid-semester ratings than they in fact received. For McKeachie, two forms of feedback were tried, printed reports of mid-semester ratings and personal counseling based on such reports. The first (print) group showed no change, while the second (counseling) group showed small but significant improvement on both of two overall ratings and on one of seven specific ratings; however, there was no improvement on the six other specifics. Pambookian found significant improvements only for those instructors whose initial ratings were in the mid-range.

Aleamoni (1974a), Braunstein, Klein, and Pachla (1973), and Hoyt and Howard (1977) seem to have gotten slightly more promising results. The first of these studies followed a plan that included the use of personal counseling; however, it differed from the studies cited earlier in that ratings were gathered at the end of the two different semesters, not within a single semester. The counseled instructors showed significant improvements, pre-rating to post-rating, on two of six rating sub-scales, although their last (or post) ratings were not significantly higher than those of the uncounseled control group instructors.

Braunstein, Klein, and Pachla randomly assigned participating instructors to experimental and control conditions, the first group receiving the summarized results of the mid-semester ratings and the second group having their results withheld. The investigators also had the first group record their expectations of the ratings they would receive. For their purposes, a change in one's ratings, pre

to post, was defined as at least one scale point. They found that 65 percent of the first group's changes were positive and 35 percent negative, while exactly the opposite was true for the second, or control, group. The difference in the frequency of these changes was statistically significant. In addition, where the experimental group instructors had overestimated the level of their initial ratings, all 13 instructors showed positive changes in subsequent ratings, whereas only four of 18 underestimates showed such change. This difference was also statistically significant.

Finally, Hoyt and Howard present two kinds of pertinent data. They analyzed earliest and latest ratings of 263 Kansas State faculty who had used the University's IDEA rating system during the period 1969 to 1972. Of 15 rated characteristics, 13 showed statistically significant increases. Secondly, they analyzed the ratings of 348 faculty who had had much, some, or no contact with the University's educational development office, an office which provides consultation services to faculty, often centered around the individual's ratings. On 14 of 18 mean ratings, the three faculty groups differed significantly and consistently in the predicted direction.

We suggest that the jury is still out and that the question of whether ratings can serve to improve instructional performance is not yet decided. Some statistically significant changes in ratings have been found, yet these have not been large, in absolute terms. It remains to be seen whether refined procedures can yield more promising results. If they do, it seems likely that they will benefit from some refinement of the diagnostic sensitivity of the rating instruments used, from considering effects over a longer term than has been customary in earlier studies, and from concerted intervention to assist faculty in interpreting and using diagnostic ratings.

Concluding Comments

In closing, an admission: At the outset, the writer intended not to present yet another review and critique of the student rating literature. None seemed needed. But as writing progressed, that intention eroded noticeably. It became impossible to discuss ratings and their uses without frequent reference to empirical work and to interpretations which others have offered. Nevertheless, the result has not been a review that parallels those published and cited earlier. To the extent that this is so, the original intention has held, and interested readers will have good reason to examine those reviews and their primary sources. Those who do are likely to conclude, as we do, that the elusive coefficients which have played hide-and-seek with many investigators are one more indication that effective teaching consists of most of those things that earlier observers have suggested, even including those that appear mutually exclusive. And in all its aspects, effective teaching remains an inviting and worthy phenomenon to study, especially if one has stamina and a high tolerance for frustration.

REFERENCES

Aleamoni, L. The usefulness of student evaluations in improving college teaching (Research Report #356). Urbana: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1974.

Aleamoni, L. Typical faculty concerns about student evaluation of instruction. Paper prepared for presentation at the Symposium on Methods of Improving University Teaching, The Technion-Israel Institute of Technology, Haifa, March 1974.

Astin, A. & Lee, C.B.T. Current practices in the evaluation and training of college teachers. In C. B. T. Lee (Ed.), Improving college teaching. Washington, D.C.: American Council on Education, 1967.

Bassin, W. M. A note on the biases in students' evaluations of instructors. Journal of Experimental Education, 1974, 43, 16-17.

Bejar, I.I. A survey of selected administrative practices supporting student evaluation of instructional programs. Research in Higher Education, 1975, 3, 77-86.

Borgatta, E. Sidesteps toward a nonspecial theory. Psychological Review, 1954, 61, 343-352.

Borgatta, E. Student ratings of faculty. AAUP Bulletin, 1970, 56, 6-7.

Braunstein, D., Klein, G., & Pachla, M. Feedback expectancy and shifts in student ratings of college faculty. Journal of Applied Psychology, 1973, 58, 254-258.

Centra, J. A. Evaluating teachers. Letter in Change, 1973, 5, 7.

Centra, J. A. Effectiveness of student feedback in modifying college instruction. Journal of Educational Psychology, 1973, 65, 395-401.

Centra, J. A. Student ratings of instruction and their relationship to student learning. American Educational Research Journal, 1977, 14, 17-24.

Costin, F., Greenough, W., & Menges, R. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511-535.

Doyle, K. O., Jr. Student evaluation of instruction. Lexington, MA: D. C. Heath & Co., 1975.

Drucker, A. J. & Remmers, H. H. Do alumni and students differ in their attitudes toward instructors? Journal of Educational Psychology, 1951, 42, 129-143.

Eble, K. Professors as teachers. San Francisco: Jossey-Bass, 1972.

Eble, K. Letter in The American Scholar, 1972, 41, 494-495.

Frey, P. Student ratings of teaching: Validity of several rating factors. Science, 1973, 182, 83-85.

Frey, P. The ongoing debate: Student evaluations of teaching. Change, 1974, 6, 47-48; 64.

Gage, N. L. The appraisal of college teaching. Journal of Higher Education, 1961, 32, 17-22.

Gessner, P. Evaluation of instruction. Science, 1973, 180, 560-570.

Guthrie, E. R. Measuring student opinion of teachers. School and Society, 1927, 25, 175-176.

Guthrie, E. R. The evaluation of teaching: A progress report. Seattle: University of Washington, 1954.

Heilman, J. D. & Armentrout, W. D. The ratings of college teachers on ten traits by their students. Journal of Educational Psychology, 1936, 27, 197-216.

Hildebrand, M., Wilson, R. C., & Dienst, E. R. Evaluating university teaching. Berkeley: University of California Press, 1971.

Hoyt, D. & Howard, G. The evaluation of faculty development programs. Kansas State University Research Report #39. Office of Education Research, 1977.

Hudleson, E. The validity of student ratings of instruction. School and Society, 1951, 73, 265-266.

Kerlinger, F. Student evaluations of university professors. School and Society, 1971, 99, 353-355.

Kossoff, E. Evaluating college professors by "scientific" methods. The American Scholar, 1971-72, pp. 79-93.

Kulik, J. A. & McKeachie, W. J. The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), Review of research in education (Vol. 3). Itasca, IL: Peacock, 1975.

Lovell, G. D. & Haner, C. F. Forced-choice applied to college faculty rating. Educational and Psychological Measurement, 1955, 15, 291-304.

McKeachie, W. J. Student ratings of faculty. AAUP Bulletin, 1969, 55, 439-444.

McKeachie, W. J. Use of student ratings in evaluation of college teaching. National Institute of Education, Grant No. NE-G-00-3-0110 Final Report, March 1975.

McKeachie, W. J., Lin, Y-G., & Mann, W. Student ratings of teacher effectiveness: Validity studies. American Educational Research Journal, 1971, 8, 435-445.

Maslow, A. H. & Zimmerman, W. College teaching ability, scholarly activity and personality. Journal of Educational Psychology, 1956, 47, 185-189.

Menges, R. J. Letter in Change, 1973, 5, 7.

National Project III. Evaluation of college teachers (Criteria Report No. VI). Ann Arbor: Center for Research on Learning and Teaching, University of Michigan, 1977.

Pambookian, H. S. Initial level of student evaluation of instruction as a source of instructor change after feedback. Journal of Educational Psychology, 1974, 66, 52-56.

Powell, R. Faculty rating scale validity: The selling of a myth. College English, 1978, 39, 616-629.

Remmers, H. H. The relationship between students' marks and student attitude toward instructors. School & Society, 1928, 28, 759-760.

Remmers, H. H., Shock, N. W., & Kelly, E. L. An empirical study of the Spearman-Brown formula as applied to the Purdue rating scale. Journal of Educational Psychology, 1927, 18, 187-195.

Remmers, H. H. Rating methods in research on teaching. Handbook of research on teaching (Chapter 7). American Educational Research Association. Chicago: Rand, McNally & Co., 1963.

Rodin, M. Can students evaluate good teaching? Change, 1973, 5, 66-67; 80.

Rodin, M. & Rodin, B. Student evaluations of teachers: Students rate most highly instructors from whom they learn the least. Science, 1972, 177, 1164-1166.

Seibert, W. F. National Project III: Elevating the importance of teaching (Fund Associate's Final Report to Fund for the Improvement of Postsecondary Education, Grant No. G007502063). West Lafayette, IN: Measurement and Research Center, Purdue University, 1977.

Seldin, P. How colleges evaluate professors. New York: Blythe-Pennington, Ltd., 1975.

Spencer, R. E. & Aleamoni, L. M. A student course evaluation questionnaire. Journal of Educational Measurement, 1970. 7, 209-210.

Sullivan, A. & Skanes, G. Validity of student evaluation of teaching and the characteristics of successful instructors. Journal of Educational Psychology, 1974, 66, 584-590.

Vattano, F. J. Letter in The American Scholar, 1972, 41, 495-496.

CHAPTER NINE

PROGRAM REVIEW AND INSTRUCTIONAL DEVELOPMENT

Robert G. Arns* and William Poland
The Ohio State University

In recent years, efforts have increased to create a more rational basis for decision-making in higher education. Universities have become substantially larger and more complex over the last three decades. During this period, they have often depended upon growth for improving current activities and for developing new directions. Now that the great period of growth in size has come to an end, improvement and change must come through choice. Wise choice demands a basis of sophisticated information.

Institutions of higher educatior are not only learning how to change without growth, but they are also experiencing a period in which knowledge and the ways of discovering, organizing, and transmitting knowledge are advancing significantly. This generates an increased need for orderly change at a time when resources are more or less stable and economic conditions--particularly the inflation of recent years--have further reduced the degree of flexibility available. "Planning" has become a key word--sometimes a euphemism for desperate flailing attempts to cope with a deteriorating situation. More often than not, first attempts have been flawed; however, now there is gradually emerging, from past mistakes and the patience of many, an understanding of the major requirements of a rational basis for decisions and of the pitfalls to be avoided.

I. The Role of Program Review in Comprehensive Planning

Successful comprehensive planning embodies some key features. First, it must take place at all levels within the institution with

---

*Dr. Arns is now Vice President for Academic Affairs at the University of Vermont.

the commitment not only of presidents and deans but also of faculty, staff, and students. Second, it must be a rational and iterative process in which the understanding of all grows through the repeated testing and refinement of ideas. Third, genuine change must result from the planning. Finally, the plans must outlast the persons who made them, and key decisions must continue to be based on these plans.

Successful planning also embraces time scales which are both short and long. It establishes principles which will guide activities such as day-to-day decisions, annual budgets, and long-range projections.

It is not unusual to encounter planning processes which project resource requirements, fund sources, enrollments, etc. forward for a one- to five-year period to establish the basis for daily incremental decisions and to project long-range trends. Processes of this type, which result in what we shall call a "term plan," do not lead to a fully rational basis for choice unless they are accompanied by (1) the establishment of an institutional context and (2) a system for periodic review of existing programs.

The expression "institutional context" is meant to include an established framework in which planning takes place. It typically involves items such as a mission statement, an inventory of goals, a statement of planning assumptions (constraints), an expression of areas of emphasis (principles used to select certain programs and goals for preferential support), and a set of priorities to link the areas of emphasis to specific programs. While the "institutional context" should continue to develop over time, at any one time it is important to have a focus (or foci) from which planning proceeds. No university can now afford to develop willy-nilly as a simple aggregation of independent fiefs.

The term "program review" is meant to convey a searching, comprehensive evaluation of an existing program. Depending on the "depth" and "comprehensiveness" of the review and the breadth of definition of "program," the term can mean different things to different people (Clark, Note 1). For purposes of the current analyses, the question of how broadly "program" is defined is not crucial. The review schemes described in the following section generally involve a broad definition, for example, "the (total) Chemistry program," as contrasted with the narrower "the M. S. program in Chemistry." We will also assume a review scheme which probes deeply into the value, quality, and resources of all activities of the program under study and fully examines relationships with other programs. While program review generally involves all teaching, research, and service functions of the unit under study, the present discussion will focus particularly on instructional aspects and describe the use of program review as a tool for exploring the relationship between intended learning objectives and the instructional designs intended to produce those outcomes.

Each of the three prototypical planning concepts described--the term plan, the institutional context, and program review--has an important role to play in comprehensive planning. The term plan tends to deal with incremental and decremental issues as they relate to existing programs. The institutional context can provide a sense of direction and purpose through which individual programs can find their appropriate role. Program review, finally, looks at the totality of an existing program as well as each program component and raises questions regarding value, quality, and the allocation of resources. It is in the comprehensive review of teaching aspects that the unique skills of the instructional developer must be brought to bear if substantive planning leading to improvement of instruction is to result.

II. Program Review in C.I.C. Institutions

While each of the Big Ten universities in the C.I.C. is sufficiently large to have been involved in most aspects of program review at one time or another, we will limit our attention to Illinois, Michigan, Minnesota, and Ohio State. Each of these four institutions has a comprehensive process for review of academic units which can serve to illustrate the patterns employed.

The University of Illinois (McIntyre, Weir, & Hake, Note 2): In 1972 a Study Committee on Program Evaluation (SCOPE) recommended guidelines for a review system to be guided by a Council on Program Evaluation (COPE). Reviews began later that year and were, for the most part, reviews of individual departments which began with a structured self-study in which the unit responded to a lengthy questionnaire. This self-study phase was followed by an analysis by a "task group" (at least seven faculty members including three from the unit under study) of (1) unit operations and practices, (2) leadership, (3) academic staff and affirmative action, (4) quality of life in the unit, (5) graduate and undergraduate programs, (6) public service and continuing education, (7) non-academic staffing, and (8) space assigned to the unit. On some occasions, at the option of the task groups, discipline experts were brought to campus to assist in the review process.

Some 18 units were reviewed in this manner during 1973 and 1974 resulting in action reports from COPE containing recommendations from each study. In each case, the "action report" included the main thrust of the findings of the task group and those recommendations that the Council accepted as both feasible and important. By 1975-76, COPE's third year of operation, the evaluation plans reflected increasing emphasis on self-study, with the subsequent assignment of task groups for more detailed study--an increasingly infrequent phenomenon.

Critics of the COPE project have complained that the self-study documents are not sufficiently flexible to be applicable to all units under study and that little change has resulted from the effort expended in the review process. Advocates believe that the process has made significant contributions to the understanding of

all involved and that, when the first round of reviews is completed, the institution will be routinely collecting information which will make it possible to decide when critical review of a particular unit is needed.

The assertion by some that the benefits are not sufficient to justify the time spent has not been unique to the Illinois experience. It is difficult to distinguish between situations in which the improvements achieved were truly small and those in which the expectations of some participants were unreasonably large at the outset. Moreover, some of the most valuable outcomes may remain unapparent because, in the course of review, changes occur and problems are solved which cannot clearly be claimed as a beneficial result of the process, even though they might not have otherwise resulted.

The University of Michigan (Ericksen & Lelong, Note 3): In 1975 the University of Michigan initiated an evaluation planning project consisting of a three-year, three-phase effort. The first phase was devoted to the articulation of operating objectives by each of the University's schools and colleges, as well as subsidiary academic units.

As originally designed, phase II was to be devoted to a careful self-evaluation by each unit, in light of the objectives defined in Phase I. Phase III was to be devoted to the formulation of planning memoranda which would represent an understanding between the academic vice president and each dean with respect to objectives and plans for that college for the next five years.

As a result of experience with the initial phase, in which the need for an iterative approach was recognized, some of the evaluative activities were moved beyond the second phase and some planning activities were incorporated into this second phase. Since the project is not scheduled for completion until August, 1979, it is premature to judge its effectiveness. It is anticipated, however, that the development of "memoranda of understanding" will contribute significantly to the potential of the project for contributing to effective planning.

The University of Minnesota (Linck & Zimmerman, Note 4): The program review system at the University of Minnesota involves a somewhat different protocol for each college. It typically includes a self-study phase, an external review by scholars from other universities, and an "internal review" involving a group of peers from within the University who are responsible for merging the results of the self-study and external review and developing a final report about the unit under study. The typical "program" is an entire department.

In the Minnesota review system, it was found that one of the most important phases consisted of "closure meetings" with the dean of the college, the dean of the graduate school, and the academic vice president. At this stage, the various parties come to terms

with relating budgetary considerations to the recommendations made by the reviewers. It was also noted that this part of the program review process establishes the credibility of the entire undertaking.

Those involved in the Minnesota reviews believe them to have been quite effective in most cases. At the same time, they recognize the substantial investment of time required and, in a few cases, some dissatisfaction with the extent of change effected.

The Ohio State University: In the autumn of 1976 after three years of studying our previous self-evaluation efforts and studying program review systems at other universities, the first program reviews began at The Ohio State University. In the following description, the reader will note major features drawn from the experience at the three institutions described above.

The primary purpose of the reviews has been program improvement, based on recognition of a continuing need to articulate goals; to test effectiveness in meeting these goals; to measure the quality, relevance, and potential of activities; and to determine the extent to which resources and the organization and alignment of academic and administrative units serve the University's mission.

While the review process has required, as background, certain descriptive information, including an understanding of the goals, activities, budgets, student characteristics, faculty characteristics, etc. of each program, emphasis has been on three basic issues: (1) Value: assessment of the nature and importance of a program's goals as they relate to the needs and goals of students, of the University, and of society. (2) Quality: assessment of the extent to which a program achieves its goals and of its strengths and weaknesses. (3) Effective use of resources: assessment of the appropriateness of the allocation and organization of human, fiscal, and physical resources to and within the program in order to decide whether and how to modify, expand, contract, initiate, or eliminate program components.

While reviews are tailored to the needs of the individual program (for example, the definition of "program" is not restricted to existing fiscal or organizational units), the following features are commonly included: (1) Self-study: preparation and interpretation of descriptive and evaluative background in terms of the issues cited above. (2) External review: a structured two-day to three-day visit by scholars from nationally recognized institutions. (3) Coordination/communication: a separate internal committee of peers to coordinate the activities and facilitate both feedback to the various parties to the review (provost, deans, program participants) and communication within the University community. (4) Closure: development of a plan of action, in the form of a memorandum of understanding, and its implementation.

Although the system is still evolving and too young to yield definitive evaluation, the results to date have been satisfying,

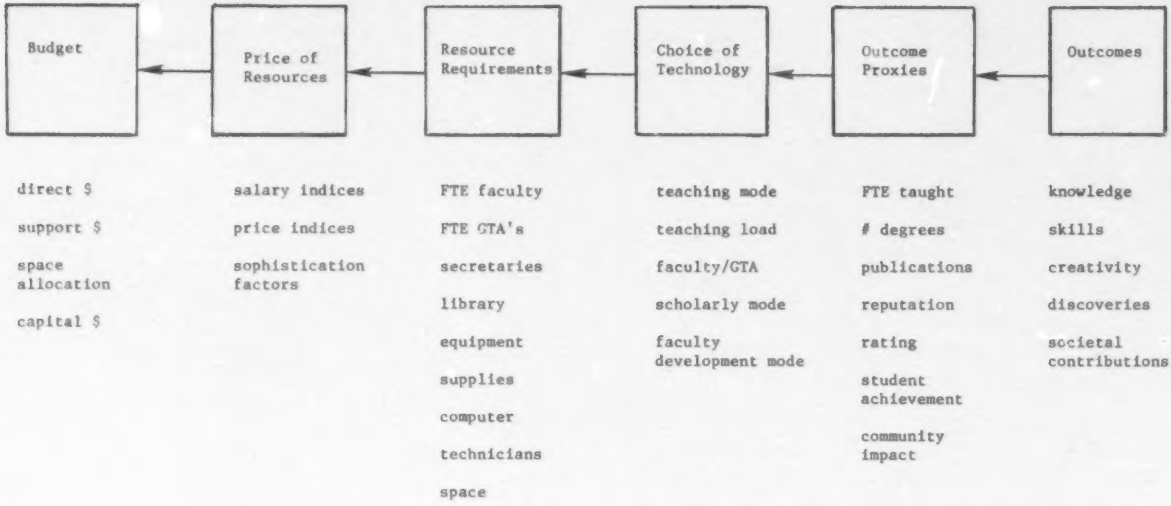


Fig. 1. Stylized representation of a stepwise analytical model for determining the resource requirements needed to produce particular outcomes. Beginning at the right with a specification of general goals, the model uses five successive steps to infer the budget required to produce these general goals. The "key words" shown under each box do not represent an exhaustive list but rather are intended to illustrate the kinds of considerations which might be appropriate to each step.

despite the large investments of time required. There appears to be no way of avoiding the effort required by this process if there is to be genuinely substantive evaluation and a rational basis for decisions.

III. Analysis of Program Elements

In addressing questions of value, quality, and effective use of resources, it is helpful to have an analytical framework as an aid to teasing out the critical questions. We have found it useful to subject each program component to a stepwise analytical process in which general outcomes form the starting point, with the resources required to produce these outcomes being the desired end point. This is represented in a stylized fashion in Figure 1. The starting point, at the extreme right of the diagram, is the enumeration of the general outcomes which the program strives to achieve. Since this initial list usually involves abstract concepts which defy measurement, it is necessary to specify, as the next step, a set of outcome proxies--indicators of performance which can provide a measurable approximation to the outcomes.

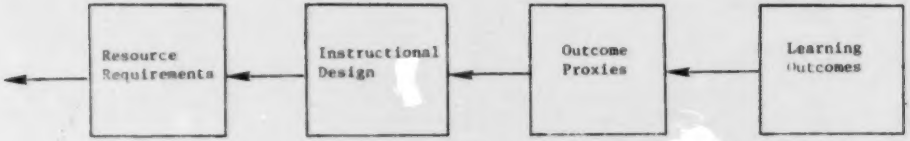
After these proxies (indicators) are specified, the next step requires selection of a set of technologies capable of producing the proxies. The word "technology" in this third list is used in the broadest sense to indicate the "mode of operation"--the "design" for achieving the outcomes.

Under the "Choice of Technology" heading, for example, the term "teaching mode" is intended to encompass notions like class size, class type (lecture, discussion, seminar, laboratory, field experience), use of computer-based learning, etc. As might be expected, "teaching load" means x number of y-credit-hour courses per academic year. The words "scholarly mode" hold a place for answers to questions like, How does the discipline develop? How does it renew itself? Similarly, the words "faculty development mode" hold a place for answers to questions like, How do faculty members in the discipline renew themselves? Do they work as individual scholars or as members of a research team? Do they work at the institution or must they travel to a distant library or laboratory?

The development of the fourth list, a list of the resources necessary to support the various "technologies," then follows with features as illustrated in Figure 1.

The fifth list, again requiring a move to the left, attaches a unit price to each of the resource requirements. This fifth list also provides an opportunity to include information about incremental costs that are, for example, due to differing inflationary trends of the various resources.

An additional consideration at this juncture has been labeled "sophistication factor" in Figure 1 and is intended to include unusual incremental cost increases/decreases tied to sudden changes



teaching mode
teaching load
faculty/GTA

FTE taught
degrees earned
number of majors advised
student achievement
student attitudes

knowledge
capacities of intellect
skills
values
personal growth

Fig. 2. Stylized representation of the "front end" of a program analysis as applied to an instructional component of a program. The items shown under each box are illustrative rather than exhaustive. The key words "teaching mode," for example, include items like course size, class type (lecture, recitation, seminar, laboratory, field experience, independent study, etc.) and media role (computer based learning, instructional television, audio-visual aids, etc.)

in technology, for example, a quantum step in costs due to introduction of on-line computer technology to a research process or due to "mainstreaming" of handicapped students.

The final step in the process is the production of a budget for the program.

For the sake of discussion, let us assume that the "program" under study represents a typical department in a college of arts and sciences with the usual range of undergraduate and graduate teaching, research, and service responsibilities. As a first approximation, it may be helpful to break such a program into a half dozen components: (1) service teaching (mainly freshman and sophomore level), (2) instruction for majors (mainly at the junior and senior level), (3) graduate "classroom" instruction (including seminars), (4) direction of theses and dissertations, (5) service responsibilities, and (6) faculty scholarship. Each of these components of the program can then be viewed in terms of the proposed model, i.e., the inferential process can proceed from right to left without reference to the other components. The individual components are not totally separable, however, and the interrelationships must be kept in mind as the analysis proceeds.

The process of describing intended outcomes and inferring resource requirements need not proceed to full quantification or to completion in order to provide information which will be helpful to those interested in understanding and improving a program. The greatest utility of this model probably resides in the assistance that it can provide, via the first two or three steps from right to left, in helping faculty and administrators to determine key issues and raise questions about specific program goals and activities. This will be illustrated in Section V.

For purposes of the present exposition, we will limit our attention to instructional programs and to the first two steps in the process, namely (1) the development of a set of proxies to represent the general outcomes intended and (2) the analysis of the technology used to produce the outcomes. When referring to teaching components, the outcomes are appropriately labeled "learning outcomes" and the "choice of technology" becomes "instructional design" (See Figure 2).

IV. Learning Outcomes and Their Proxies

At the outset, it is necessary to recognize that the goals of an instructional program are usually abstract and idealized. Although concepts like knowledge, skill, creativity, and contributions to society may not lend themselves to direct assessment, they may be the starting point for any analysis. The next step requires substitution of proxies (indicators) for these general outcomes. This step, and each additional step in such modeling, represents a radical transition--a leap of inference from one or more items on a given list to one or more corresponding items on the next list to the left--and an act of faith that the

substitution does not distort the original goals. It is essential that the total picture be kept in mind and that there be constant questioning of the sufficiency of the information retained as the process moves step-by-step to the left.

We will not attempt here to review the ample literature on methods for defining and assessing outcome proxies (NCHEMS, Note 5). This literature contains numerous references to "setting goals," "translating goals into measurable objectives," "data collection instruments," "student achievement measures," "survey instruments," "community impact assessment," etc. While these research and development efforts have provided information and insights which may be useful in particular situations, the techniques developed to date have not been widely applied or accepted.

It is important to note that the faculty members, whose commitment is essential to the successful review of their program, often begin the endeavor with a cultural bias peculiar to their discipline. Most of them have not studied its pedagogy or been involved in tests of instructional methods. They tend to have drawn their own teaching styles from their teachers. They often bring to the study of their instructional activities the style of thought and research characteristic of their own scholarly work. This can lead to semantic confusion and, initially, to a limited approach to a particular problem. To cite some cruel stereotypes, the professor of mathematics, in analyzing a pedagogical issue, may state a theorem and set out to prove it, while the professor of an experimental science may confront a similar issue by attempting to design a tightly-controlled experiment, and the professor of education may set out to survey student opinion. Those who assist the review process must build upon the natural strengths and styles that faculty members bring to the task. They must assist them in broadening their horizons and in using a multiplicity of methods to seek answers.

Faculty members also often suffer discomfort in attempts to be specific about goals or performance measurement, even where semantic confusion and cultural differences do not play a role. There is an inherent safety in imprecision and a generally valid sense that completeness cannot be achieved; this must be accepted.

Rather than attempting to design universal measuring instruments or to achieve broad consensus on an appropriate set of goals or proxies for a given program, it is better to make and test some tentative choices at an early stage, recognizing that it will probably be necessary to repeat the process until a sufficient set has been identified. Above all, the proxies chosen should be tailored to the situation and none should be taken very seriously until the program participants and the other parties to the review are comfortable that they represent an acceptable approximation to the outcomes sought.

In order to take the next step in the process, the proxies need to represent not only measures of quality but also measures of

quantity. One typical quantitative measure is the number of full-time-equivalent students taught per unit time. Another is the number of degrees earned per year. Measures of quality are more difficult to establish and more difficult to interpret. Examples which come to mind include average scores on standard examinations, fraction of students admitted to advanced study (for certain types of programs), fraction of students in appropriate careers, rate at which students advance after graduation, student satisfaction while taking courses and after graduation, and extent to which alumni are recognized leaders. None of these outcome measures of quality is susceptible to full interpretation in the absence of comparative data, either from other institutions or from other cohort groups within the institution. But with a sufficiently wide spectrum of such information and comparative data to permit interpretation, it is possible to construct a set of proxies of sufficient credibility to test current instructional design or changes in instructional design.

V. Analysis of the Instructional Design

Having established a set of outcome proxies, however tentative, the next step is to study the processes by which these proxies are generated--the instructional design. By "instructional design," we mean all aspects of the way in which students are taught and learn through the curriculum under study, including such things as the selection of course materials; the use of technological aids; class type (lecture, discussion, seminar, laboratory, field experience, independent study, etc.); class size; who teaches, e.g., faculty and/or graduate teaching associates (GTAs); teaching load; course content; and feedback and evaluation (grading) methods. Presumably matters of this type are addressed course-by-course or, better, are addressed for groups of related courses, e.g., all general education courses in history intended for lower division students.

At this stage of the review, probing of the instructional design brings forth such questions as:

> Considering the varying class sizes in our introductory courses, should we eliminate certain undersubscribed courses and increase enrollment limits in the remaining courses to accommodate the same total number of students? If this is done, what will be the effect on the value and quality of this part of the curriculum and the effect on the quality of the courses which remain? Should we change to a lecture and recitation format using faculty and GTAs, respectively, in place of the current arrangement in which faculty and GTAs teach different sections, each with full responsibility?
>
> Can we design an effective introductory science course for non-scientists without a laboratory component? What abilities (objectives) will be foregone by omitting the laboratory? Can some of these be provided in other ways?

> Can students acquire the same abilities at this stage of their development with different course content? With a different mix of human instruction and instructional materials or instructional technology?
>
> Can the goals of our master's program be achieved if the thesis requirement is dropped?
>
> Should we adopt competency-based learning for some of our courses? Which courses?
>
> Are we trying to do too much in our curriculum? Could we be more selective in the content and abilities we profess to teach and by so doing both lower teaching loads and enhance learning?

Although the foregoing questions are, in a sense, relevant to single components of the program, some of them may contain implications for other components. The last of the above questions is such an example. In examining the possibility of reduced teaching loads, one may find an opportunity for increased faculty scholarship. Other questions which may involve more than one component of the program are, for example:

> Is faculty effort appropriately deployed among graduate seminars, courses for undergraduate majors, and introductory service courses?
>
> There are no prescribed courses in our major program. Each student's major is individually designed with the result that the advising load is heavy and there is considerable variation in the major programs. Should we have a required core of courses for each major? Can we, at one and the same time, improve the quality of the undergraduate program in this way and reduce the faculty advising load?
>
> Given that the faculty of this department is of appropriate size and given that, in view of the job market and other considerations, we have too many Ph.D. students (because of our dependence on graduate teaching fellows as teachers in a key introductory course), can instruction be delivered, at least in part, effectively and economically in some other way so that the size of the graduate program can be reduced? (The answer was affirmative. Using a combination of lectures with computer-based learning, student satisfaction and learning were improved. Because excess capacity was available on an existing computer, this was accomplished without additional cost.)

Faced with questions of this type, which arise in introspective review activities, the faculty of a program frequently turns to pedagogical literature of its discipline and to the advice of peers at other institutions. The quality of pedagogical literature varies considerably from discipline to discipline. Faculty members or

instructional development staff members who are familiar with the pedagogical literature of several different disciplines can often find helpful information from work in allied fields or from work on similar problems.

Comparative information from similar programs at other universities can also provide a useful starting point (See Chap. 3). Chairmen of certain departments of the C.I.C. institutions, for example, meet annually to share information on a variety of topics. The authors have examined comparative data for this group of institutions in six different disciplines. Similarities and differences in instructional design were identified with the help of representatives of each discipline. It was observed that there are many similarities within each discipline studied and that the differences helped to raise fundamental questions of strategy. Even when similar departments show similar instructional patterns, it is important to ask whether the choice of instructional design is genuinely characteristic and essential or merely historical and traditional. This last question, if squarely faced, counsels reference to the general literature of educational research, which may provide guidance in some instances. This kind of question also often calls for the design of evaluation schemes to test the premises on which the instructional design is based.

VI. Role of the Instructional Development Staff

Approaching the review of an academic program through the processes described in the foregoing sections makes it possible for those involved to raise questions which might otherwise be missed. Some of these questions are very general and profound, and some are very specific. The processes described do not provide answers. They improve the chances that answers will be sought, and they represent an important role for the instructional development staff in assisting the faculty. The steps outlined--(1) identification of general learning outcomes (goals), (2) establishment of proxies (indicators) for these outcomes, and (3) analysis of the instructional design being used--begin with tentative choices and admittedly imperfect understanding; they are repeated and refined until appropriate courses of action are found and until mechanisms are in place for evaluating the results of these courses of action.

The instructional developer, as a participant in this process, should have:

1. Familiarity with and the ability to apply the general literature of educational research, including the study of outcomes, outcome assessment, and instructional strategy. This includes the ability to cope with the implied myths in the example questions in Section V--such as "small classes ensure better learning" or "greater flexibility means better curricula."

2. Access to the pedagogical literature of various disciplinary specialties. In this and in the previous item, the ERIC data bases can be especially helpful.

3. Access to comparative information from similar programs at other institutions and/or within the same institution.

4. The ability to design credible evaluation schemes for situations ranging from an entire curriculum to an individual course strategy.

The instructional developer need not be involved in every aspect of these processes. In the institutions studied the instructional development staff has not been heavily involved in the design and operation of the overall program review processes. Instructional development staff members have participated in activities such as providing workshops on instructional design issues and constructing and testing assessment instruments for use on a university-wide basis in program reviews. The most important and appropriate role for the instructional developer in individual program reviews is in assisting the faculty to define and solve specific problems or to answer specific questions of the type illustrated in Section V.

The instructional developer must not be viewed by the parties to a program review as an evaluator. He or she should function as a resource to facilitate the efforts of the faculty to understand and improve program components. If the instructional developer is to play a useful and catalytic role, he or she must be accepted as part of the departmental team and have the confidence of the faculty. Involvement with individual program reviews should therefore be by invitation, and at a sufficiently early stage to enable the "outsider" to establish credibility and legitimacy in the eyes of the faculty.

Planning and program review are intended to yield information which will provide a rational basis for choice. If there is to be a substantive leap in accomplishment in these activities, it is necessary to get to the heart of a program and examine fundamental assumptions. Program review provides at one and the same time an occasion for such probing and a special challenge to those who assist faculty to improve their programs. The instructional development staff has an important role to play in the process.

REFERENCE NOTES

1. See, for example, Clark, Mary Jo. Program Review Practices of University Departments. G. R. E. Board Research Report GREB No. 75-5aR, July, 1977.

2. McIntyre, Charles J., Director, Office of Instructional Resources; Weir, Morton K., Vice Chancellor for Academic Affairs. Personal communications. Also based on reports and memoranda prepared by Harold W. Hake, Associate Vice Chancellor for Academic Affairs and Chairman of the Council on Program Evaluation, University of Illinois.

3. Ericksen, Stanford C., Center for Research on Learning and Teaching; Lelong, Donald C., Director, Office of Academic Planning and Analysis, The Unive sity of Michigan. Personal communications, 1978.

4. Linck, Albert J., Associate Vice President for Academic Affairs, & Zimmerman, Kenneth, Assistant Dean of the Graduate School, University of Minnesota. Personal communications.

5. See, for example, National Center for Higher Education Management Systems at WICHE. Outcome Measures and Procedures Manual (Tech. Report 70). Boulder, 1975. Also various articles in the reports of the annual fora of the Association for Institutional Research.

CHAPTER TEN

THE SYSTEMATIC INVOLVEMENT OF STUDENTS IN PROGRAM EVALUATION

Charles J. McIntyre
University of Illinois--Urbana-Champaign

In the preceding chapter Drs. Arns and Poland define program evaluation and give some examples of program evaluation presently occurring in various universities. They predict that the years ahead will bring an increase in the frequency of its application.

As the chapter makes clear, there are various models of program evaluation. For the purpose of this discussion, I will refer to the model used at the University of Illinois--Urbana-Champaign (UIUC). Characteristically, program evaluation here has involved the evaluation of an academic department in all of its aspects--scholarship, teaching (both graduate and undergraduate), continuing education, disciplinary leadership, relations with students other than formal teaching (such as advising and helping further their careers), administration of the unit, costs, morale of faculty, students and other employees, and frequently other factors. Typically, many statistical data are available bearing on many of these matters. Typically, too, a department has (or can create) well thought-out statements on many of them. But, finally, some other inquiry and judgment may have to enter the process.

At the University of Illinois (as well as elsewhere) committees have been empaneled to study these matters in detail. At UIUC these committees report to a Council on Program Evaluation. Committees established for evaluative purposes generally include faculty both from within and outside of the department being evaluated and usually one or two students majoring in the department.

Another component of program evaluation at the University of Illinois, which seems to be absent from evaluations at other institutions--or at least less systematically investigated has been that of comprehensive student input. This was begun here because it

was believed that students who had experienced important aspects of a department's program could make important evaluative judgments about it. Of course, these judgments must be gathered, analyzed, summarized, and disseminated in a way both responsible and consonant with accepted psychometric principles.

I. The Development of the Program Evaluation Survey (PES) (Note 1)

To obtain comprehensive student input, staff of the Office of Instructional Resources developed an instrument which came to be called the Program Evaluation Survey (PES). The information obtained from the students with the PES was to serve University-level administrators and the committee evaluators who could use the information in identifying relative strengths and weaknesses within each department, so that program improvements could be implemented if needed.

In constructing the survey the following guidelines were held uppermost: (a) The information asked of students majoring in the program would be the kind of information that they could appropriately provide; (b) one instrument could be administered to all students regardless of undergraduate or graduate status or academic department affiliation; (c) students would be able to complete the instruments in less than fifteen minutes; (d) the results could be presented in a format that would be understood easily by faculty and administrators not familiar with measurement terminology and symbolism; and (e) the instrument could be shown to be reliable and valid.

The current version of the PES is the third developed over a three-year period. The original form consisted of forty-one items and was used in the evaluation of five academic departments. The first revision consisted of thirty-three items plus four open-ended questions administered in twenty-two departments. Subsequently, minor revisions were made in this form so that the current version consists of twenty-four bipolar adjective items plus the same four open-ended questions. In each instance, revisions were based mainly on two criteria: (a) item statistics such as adequate reliability and discrimination; and (b) comments from users.

The current version of the PES contains six biographical items (sex, class-level, accumulative GPA, transfer status, graduation status, and reason for choosing the major). The twenty-four substantive items ask students about the level of difficulty and structure of their major programs, admission standards, their attitude toward their major program, and their satisfaction with eighteen programmatic areas of their department, such as integration of courses, academic contact with faculty, quality of vocational guidance, and value of the program.

The Program Evaluation Survey is administered only to majors in departments being evaluated. The departments are responsible for the administration, which typically occurs as students register.

Sample Questions from the PROGRAM EVALUATION SURVEY form, Measurement and Research Division, Office of Instructional Resources, University of Illinois.

Section I states: This "Program Evaluation Survey" was designed to elicit your opinions about the program of studies you are currently pursuing (i.e., your major). When you are answering the items please consider your reaction toward your major and your department as a whole and not specifically toward any one course or instructor. This survey is one facet of an overall departmental evaluation." In this section, students also specify class level, number of semesters on campus in the major, and cumulative GPA.

Section II states: "For each of the items below indicate your opinion of your major or your department by marking the point on the scale which most nearly expresses your opinion." Students mark a five-point scale on which the middle point indicates an opinion of "About Right." Questions are:

1. The level of difficulty of your major program. (Too Easy to Too Difficult)
2. The amount of structure in your academic program (curriculum). (Too Flexible to Too Rigid)
3. The departmental admissions standards. (Too High to Too Low)
4. The orientation of the majority of your course work. (Too General to Too Detailed)
5. Indicate your perception of the balance between departmental attention to undergraduate and graduate programs. (Undergraduate Emphasis to Graduate Emphasis)
6. Some students are more absorbed in, and enthusiastic about their majors than others. Describe your attitude toward your major program. (Do not consider "About Right" position; Indifference to Dedication)

Section III states: "Rate your level of satisfaction with your major or your department by marking one point which most nearly expresses your attitude." Following are nine of the twenty-four questions:

7. The extent to which you are challenged by your program of study.
8. The extent to which your major program provides a well-integrated set of courses.
9. The overall quality of instruction in your major department.
10. The extent to which a sound theoretical framework is developed in your major program.
11. On the average, the quality of texts and other instructional materials used by instructors in your department.
12. Classroom evaluation procedures (graded tests, papers, etc.)
13. The availability of enrichment activities offered in addition to regular classes (seminars, colloquia, social events, etc.)
14. The variety of course offerings.
15. Departmental attention to providing students with credentials for obtaining employment after graduation.

II. Analysis of the Value of the PES

In evaluating the usefulness and face validity of the instrument, extensive interviews were conducted with forty-four department heads or their representatives (Note 2). The outcomes of those interviews seem to support the following tentative conclusions.

If the purposes of a review process are both formative and summative, then a comprehensive evaluation in which information is collected from a variety of sources is desirable. For formative evaluation the departmental staff can use information about its current operation so that programmatic changes can be initiated if desired. In summative evaluation, the administrators need information for general policy decisions, resource allocations, and for rewarding, supporting, and terminating programs.

Student satisfaction, achievement, and judgments of the departmental faculty, environment, and management can be used as a part of the information base for both formative and summative evaluations. Perhaps the greatest contribution made by student evaluation is in providing information about the interpersonal, professional, and academic environment of a department. Through their daily contact with faculty and the academic program over an extended length of time, students can provide information about the degree of encouragement and support they receive and the extent to which they consider their program to be worthwhile. This type of information appears most beneficial to those most closely involved with the program operation, the department staff.

Although the purpose of the PES was to provide information to both central administrators, to assist them in making comparative judgments about departments, and to faculty, for identifying program strengths and weaknesses, the best use of the PES seemed to be for formative evaluation. There was considerable use of the PES results for this purpose. Nearly all of the departments either presented the results at faculty meetings, had select committees review the results, or made the reports available to faculty and students.

What did the departmental staff members learn from this feedback? Two of every three interviewed indicated that the students' perceptions had verified or documented their personal perceptions of student opinion, and only a few were somewhat surprised by the extent of student dissatisfaction. Departments took a variety of actions based on the feedback, resulting in some of the following actions: meetings were held with students to improve communication; committees on program evaluation were established; revisions of the curriculum incorporated student opinions; and improvements were made in academic advising.

Finally, the distinction between formative and summative evaluation becomes not all that discrete when information is used by a variety of people in a continuous review process. Similarly, departmental staffs found that the interdepartmental comparisons assisted them in better understanding their own strengths and weaknesses.

III. A Critique of the PES Instrument

In addition to interview data, extensive analyses of the instrument were run in order to try to demonstrate statistical validity. Three types of validity were examined: structural, discriminate, and criterion-related.

Structural Validity. The structural validation of the PES pertains to the identification of stable and interpretable areas of student satisfaction with the instructional and academic environment in the major's department. Do the same interpretable factors account for the responses of both undergraduates and graduates, i.e., is the same factor structure stable for various groups? On the basis of the hierarchical factor analyses performed on both the 1975 and 1976 data, the answer to the preceding question is affirmative.

According to the factor analysis, the 1976 PES instrument measures four dimensions of student satisfaction with the academic department. The same four first order factors, which are internally consistent and reasonably independent, were obtained from the responses of both undergraduates and graduates.

The first factor, Satisfaction With Mentorship, relates to faculty-student relationships. The four items measure students' satisfaction with the academic contact with faculty in and outside the classroom, faculty-student communication, quality of academic advising, and quality of vocation guidance. (Items 16, 17, 18, and 19)

The second factor, Personal Value, relates to student effort and the interest the student has in the department. The three items measure student dedication to major program, worthwhileness of the major program, and overall satisfaction with the major program. (Items 6, 23, and 24)

The third factor, Satisfaction with Instruction, represents the students' assessment of the quality of instruction received in the major department. The four items measure student satisfaction with texts, classroom evaluation procedures, major program having a sound theoretical framework, and quality of instruction. (Items 9, 10, 11, and 12)

The fourth factor, Satisfaction with Curricular Offerings, relates to the learning environment and opportunities supported by the department. The three items measure the availability of enrichment activities, integration of cours s, and variety of course offerings. (Items 8, 13, and 14)

A second order factor labeled Overall Satisfaction with the Academic Department was also obtained. Most of the PES items also reflect a pervasive departmental impression or reputation that influences the students' satisfaction with the various areas of their major program of study. This general factor may thus be representing a "halo effect."

Based on another factor analysis of the 1976 PES items and an additional set of 22 bipolar adjectives administered to majors in five departments, student satisfaction as measured by the PES appears to be dependent upon the extent of (a) departmental support and facilitation of personal and intellectual development and (b) departmental challenge. Items loading on "Satisfaction with Mentorship" reflect the supportive function of the department, while the items of the other factors reflect both support and challenge.

Scales consisting of the items with high loadings on each of the four specific factors were constructed. Coefficient alphas for each scale were generally quite high (above .70), and the intercorrelations among the scales were considerably lower (around .50). Each of the four scales of the PES are considered to measure a specific area of student satisfaction with an academic program and are interpretable for both undergraduates and graduates.

Discriminant Validity. Since the PES was to provide information upon which comparative judgments among departments and identification of weaknesses and strengths within a department can be made, analyses were conducted to assess the discriminant validity of the PES. Discriminant validity for between-department comparisons refers to the ability of the PES scales to distinguish departments on the basis of the students' responses to the items in each scale. The analyses involved the calculation of Horst reliabilities of departmental means on the 1975 PES scales (the three scales extracted in the factor analyses based on 1975 data were similar to the first three obtained from the 1976 data), which were then used to determine 95% confidence intervals of each departmental mean for each scale. These confidence intervals were then placed on a reference distribution of the scale scores of all undergraduates and graduates regardless of departmental affiliation. Based on the degree of overlap among the confidence intervals, the PES scales, particularly the scale "Satisfaction with Mentorship," were shown to possess reasonable discriminatory power for making interdepartmental comparisons. The PES is a better discriminator among undergraduate programs than graduate programs, however, partly due to the effect department size has on the reliability coefficients. Thus the PES may be most useful for evaluating large departments.

The validity of the PES for identifying within-departmental weaknesses and strengths was also investigated. Do students in a department respond differently to the PES scales so that the instrument demonstrates sufficient diagnostic value? Based on the overlap of 95% confidence intervals for the three scale means for each department, no overlap occurred between at least two of the three PES scales in nine of the 42 departmental programs (20 undergraduate and 22 graduate) reviewed in 1975. If a less stringent criterion such as a 68% confidence interval were used, many more nonoverlapping intervals would occur. These findings provide some support for the diagnostic quality of the PES, even though the overall satisfaction or rating halo factor partially limits the power of the PES as a diagnostic tool.

Criterion-related validity. The criterion-related va of the PES was investigated to determine if student satisfac with different areas of a department is related to other dep .ment characteristics regarded as important criteria for judging quality. Is student satisfaction related to characteristics of the faculty and department?

Based on the correlations between the four 1976 PES scales and 24 characteristics of faculty, student satisfaction is uncorrelated with the average number of scholarly publications, but student satisfaction with mentorship is related to the number of "other" publications. Both undergraduate and graduate satisfaction, especially with mentorship and with instruction, are positively related to membership on college committees but negatively related to membership on university committees.

However, undergraduate and graduate satisfaction are not consistently related to other faculty characteristics. Undergraduates tend to be relatively less satisfied with departments whose faculty are older and have a greater proportion of full professors, whereas graduate students are more satisfied with their faculty when the department includes a relatively large number of senior faculty. Faculty member estimates of their departmental morale is not related to graduate student satisfaction but slightly negatively related to undergraduate satisfaction.

In sum, students, especially the undergraduates, seem to base their satisfaction with their major department on the supportive and interpersonal relationships available to them within their academic department. Other correlations--or the lack thereof--are difficult to interpret.

These studies are not wholly conclusive, but as a result of the interviews with faculty users and of the statistical studies, we infer that the instrument is quite useful and appropriate for the purposes for which it was intended. We have also demonstrated to our present satisfaction the broader generalization that systematic student input as a part of program evaluation is both feasible and useful.

IV. The Value of Student-based Data.

An interesting, speculative question arises as to whether we might not be at the beginning of another controversy in evaluation similar to the one surrounding the use of student ratings in evaluating teacher-effectiveness. It was noted in Arns' paper that program evaluations which involve faculty committees, when done conscientiously and in detail, are prodigiously time consuming. Those responsible must continuously weigh whether the results are worth the cost. This does not seem to be the kind of question that can be answered in a categorical way, but will have to be evaluated anew with each new situation. It appears to us, however, that there are very likely to be times when student evaluation by means of carefully constructed questionnaires may be a larger component of

the evaluation than we would originally have thought. This is because they contain a good deal of quite useful information and can be administered, scored, and reported with relative ease and understandability.

Anyone familiar with the controversy surrounding student ratings of instruction as a means of evaluating teacher effectiveness will recognize many similar elements (Note 3): (1) Students, especially majors in the department, have a great deal of information upon which to base judgments of that department; (2) students, no matter how acute, must be unfamiliar with much of the information needed to evaluate a department; (3) hence, students probably can provide some important information but cannot provide a comprehensive picture; (4) others, mainly faculty, are needed to fill out this picture--but their time is in such short supply that it must be used sparingly; (5) student ratings can be reduced to scaleable numbers, and these can provide a kind of summary evaluation and comparison with other units; (6) having numerical data is a great convenience. However, it is absolutely imperative that those interpreting the numbers fully recognize their limitations. (It is common for people to give lip service to this warning, but frequently to ignore it.)

We note that due to considerations of economy and faculty time (and perhaps due to lack of faculty interest as well) faculty participation in judgments of their colleagues' teaching effectiveness tends to be minimal. Student evaluations, while often damned for their incompleteness, nevertheless carry the main burden in many institutions. It will be interesting to observe to what extent student evaluations of departments also become disproportionately utilized for about the same reasons.

Our present position on that issue would be about the same as it is for student evaluations, namely, that student evaluations can provide a great deal of important information but that they are generally not sufficient for the comprehensive evaluation of teaching effectiveness. Similarly, we think that student evaluations of departmental programs, while useful, are likely to be quite inadequate when used as an only source.

REFERENCE NOTES

1. Much of this chapter is taken directly from a more detailed report: J. Bessai, D. Brandenburg, L. Braskamp, S. Derry, C. McIntyre, J. Slinde, R. Smock, The Development and Use of the Program Evaluation Survey (PES): A Measure of Student Majors' Perceptions of and Satisfaction With Their Academic Department. University of Illinois--Urbana-Champaign, August, 1977, 74 pps., ERIC #HE009311, Clearinghouse on Higher Education, George Washington University, Hardcopy: $3.50, Microfiche: $.83.

2. We were not so successful in being able to interview administrative officers or the Council on Program Evaluation. All conclusions from this aspect of the study necessarily reflect mainly departmental views.

3. McIntyre, C., Seibert, W., & Owens, R. Evaluation of college teachers. Criteria VI, Center for Research on Learning and Teaching, Ann Arbor, MI, May, 1977, 6 pps.

END

DEPT. OF HEW

NAT'L INSTITUTE OF EDUCATION

ERIC

DATE FILMED

APR. 28. 1980